

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2003, 03:18:03 ; Search time 10034 Seconds

(without alignments)
11717.693 Million cell updates/sec

Title: US-10-081-408-1

Perfect score: 4040
Sequence: 1 gtcctccacccttagtc.....ccccaaaaaaaaaaaaa 4040Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_da:*
- 2: gb_hlg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_da:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_ov:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_hlg_hum:*
- 31: em_hlg_in:*
- 32: em_hlg_other:*
- 33: em_hlg_mus:*
- 34: em_hlg_pln:*
- 35: em_hlg_rtd:*
- 36: em_hlg_mam:*
- 37: em_hlg_vtc:*
- 38: em_sy:*
- 39: em_higo_hum:*
- 40: em_higo_mus:*
- 41: em_higo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4039.6	100.0	4040	6	AX329732	AX329732 Sequence
2	4039.6	100.0	4040	6	AX330309	AX330309 Sequence
3	4039.6	100.0	4040	9	HSU93447	U93447 Human Place
4	2501	61.9	2501	9	AF067406	AF067406 Homo sapi
5	2277	56.4	2277	9	AK057272	AK057272 Homo sapi
6	1933.2	47.9	3624	4	AB019242	AB019242 Bos tauru
7	1853.2	45.9	189149	2	AC046171	AC046171 Homo sapi
8	1853.2	45.9	194567	2	AC055866	AC055866 Homo sapi
9	1849	45.8	3233	9	AB050500S3	AB050502 Homo sapi
10	1787.6	44.2	2605	4	BTY15774	BTY15774 Bos tauru
11	1779.2	44.0	1794	9	AF054985	AF054985 Homo sapi
12	1761	43.6	2614	9	AB050500S1	AB050500 Homo sapi
13	1754	43.4	2664	4	BOVRA	L27218 Bos tauru
14	1754	43.4	2664	4	BOVRA	S69583 copper amin
15	1705.6	42.2	4210	10	AF115411	AF115411 Mus muscu
16	1702.4	42.1	4452	10	AF054831	AF054831 Mus muscu
17	1432.6	35.5	2611	4	AB022282	AB022282 Bos tauru
18	1359.2	33.6	2073	9	AF047485	AF047485 Homo sapi
19	1198.8	29.7	195444	2	AC025424	AC025424 Mus muscu
20	1198.8	29.7	221184	2	AL732315	AL732315 Mus muscu
21	1198.8	29.7	250353	10	AL590969	AL590969 Mouse DNA
22	1197.2	29.6	14357	10	AF078705	AF078705 Mus muscu
23	1146.6	28.4	2653	9	AF081363	AF081363 Homo sapi
24	1017.8	25.2	2572	9	D88213	D88213 Homo sapien
25	773.8	19.2	3741	9	AB012942S1	AB012942 Homo sapi
26	714.4	17.7	1051	6	AX401797	AX401797 Sequence
27	714.4	17.7	1051	6	RNU72632	U72632 Rattus norv
28	701.6	17.4	59015	2	AC067811	AC067811 Homo sapi
29	422.2	10.5	2441	6	AX330273	AX330273 Sequence
30	422.2	10.5	2441	6	AX335503	AX335503 Sequence
31	422.2	10.5	2441	9	HSU11862	U11862 Human clone
32	419	10.4	2455	9	BC014093	BC014093 Homo sapi
33	396.2	9.8	2668	10	RMABPL	X73911 R.norvegicu
34	393.2	9.7	2375	9	HUMABPA	M55602 Human amilo
35	393.2	9.7	2375	11	G18279	G18279 SWS1267 Er
36	388	9.6	2249	9	AK092514	AK092514 Homo sapi
37	381.8	9.5	2662	10	BC034215	BC034215 Mus muscu
38	380.2	9.4	2729	10	BC021880	BC021880 Mus muscu
39	355.2	8.8	2473	9	HSU11863	U11863 Human clone
40	337.2	8.3	1708	10	BC022627	BC022627 Mus muscu
41	320.4	7.9	360	6	AX017501	AX017501 Sequence
42	293	7.3	1001	9	AB050500S2	AB050501 Homo sapi
43	263.2	6.5	271	9	HUM170YACX	L77596 Homo sapien
44	259.2	6.4	991	9	AF047486	AF047486 Homo sapi
45	245.2	6.1	1721	9	AF218035	AF218035 Homo sapi

ALIGNMENTS

RESULT 1	AX329732	LOCUS	AX329732	4040 bp	DNA	linear	PAT 09-JAN-2002
DEFINITION	Sequence	241 from Patent WO0194629.					
ACCESSION	AX329732						
VERSION	AX329732.1	GI:18102710					
KEYWORDS							
SOURCE	human.						
ORGANISM	Homo sapiens						
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.						
TITLE	Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,						
	Hortigan,S., Soppet,D.R. and Weaver,Z.						
	Cancer gene determination and therapeutic screening using signature						

gene sets
Patent: WO 0194629-A 241 13-DEC-2001:
Avalon Pharmaceuticals (US)
Location/Qualifiers
FEATURES
source 1. .4040
/organism="Homo sapiens"
/db.xref="taxon:9606"
BASE COUNT 825 a 1183 c 1097 g 934 t 1 others
ORIGIN
Query Match 100.0%; Score 4039.6; DB 6; Length 4040;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4039; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCCTTCCACCTTATGACCTCCAGGATGTGACTACCGGGAACCTTAGCCAGAGTCCGGGA 60
Db 1 GTCCTTCCACCTTATGACCTCCAGGATGTGACTACCGGGAACCTTAGCCAGAGTCCGGGA 60
QY 61 GCCCCCAACCCCGTCCAGAGGACCAAGAGCCCGCTGCTGGCGTGAATACATTG 120
Db 61 GCCCCCAACCCCGTCCAGAGGACCAAGAGCCCGCTGCTGGCGTGAATACATTG 120
QY 121 CTCCTCTTGGTGAATGAGCTGTCCCTTCGTGGGAAATGAACAGAACATCTCT 180
Db 121 CTCCTCTTGGTGAATGAGCTGTCCCTTCGTGGGAAATGAACAGAACATCTCT 180
QY 181 CGTGTCTCTCATTTGCGCGCTCATACCATCTTTGCTTTGTTGTCTCTGCTGGG 240
Db 181 CGTGTCTCTCATTTGCGCGCTCATACCATCTTTGCTTTGTTGTCTCTGCTGGG 240
QY 241 CAGGGGTGAGATGGGGGTGAACCCAGCCAGCTCCCATTCCTCTGTATCTCCAG 300
Db 241 CAGGGGTGAGATGGGGGTGAACCCAGCCAGCTCCCATTCCTCTGTATCTCCAG 300
QY 301 TGCCAGCCTTGGACACACCTGCGCCAGAGCCAGCTGTTTGCAGACTGAGCCGAGAGA 360
Db 301 TGCCAGCCTTGGACACACCTGCGCCAGAGCCAGCTGTTTGCAGACTGAGCCGAGAGA 360
QY 361 GCTGAGGCTGATGAGGCTTTCTGACAGGCGGTGGGCCAGAGGCGTGTGATGACAC 420
Db 361 GCTGAGGCTGATGAGGCTTTCTGACAGGCGGTGGGCCAGAGGCGTGTGATGACAC 420
QY 421 CCAGGCGCGGCGCTGAGCAACTGTCTTCTCATGAGTGTGAGTCCCTCCCAAGGC 480
Db 421 CCAGGCGCGGCGCTGAGCAACTGTCTTCTCATGAGTGTGAGTCCCTCCCAAGGC 480
QY 481 TGACAGCCTTGCTCACTTGGACAGGGGGAGCCCCACCTGCGCGGAGGACACTGGCAT 540
Db 481 TGACAGCCTTGCTCACTTGGACAGGGGGAGCCCCACCTGCGCGGAGGACACTGGCAT 540
QY 541 CGTCTTCTTGGAGAGCAACCCCAAGCAAGTGAAGTGAAGTGTGGGCCACTGGC 600
Db 541 CGTCTTCTTGGAGAGCAACCCCAAGCAAGTGAAGTGAAGTGTGGGCCACTGGC 600
QY 601 TCACCCCTCTCATGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 660
Db 601 TCACCCCTCTCATGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 660
QY 661 CCGAGCGCCCGCTGCTTCCAAAGAGTACCTGACATAGACAGATGATCTTCAACAGAGA 720
Db 661 CCGAGCGCCCGCTGCTTCCAAAGAGTACCTGACATAGACAGATGATCTTCAACAGAGA 720
QY 721 GCTGCGCCAGGCTTGTGGGCTTCTCACCACACTGTGCTTCTCAAGAGCCGGGAGGAA 780
Db 721 GCTGCGCCAGGCTTGTGGGCTTCTCACCACACTGTGCTTCTCAAGAGCCGGGAGGAA 780
QY 781 CCGTGTGACATGACAGAGGCTCCCGTGTGTGCAATCAGGGGACCGGGCCACTGGTT 840
Db 781 CCGTGTGACATGACAGAGGCTCCCGTGTGTGCAATCAGGGGACCGGGCCACTGGTT 840
QY 841 TGGCTTCACTACAACATCTGGGGCTGGGCTTCTTCTGACCAAGTGGGCTTGGAGCT 900
Db 841 TGGCTTCACTACAACATCTGGGGCTGGGCTTCTTCTGACCAAGTGGGCTTGGAGCT 900

QY 901 GCTAGTGAACCAAGAGCCCTTGACCTTGCCCGCTGAGCTATCCAGAAAGTGTCTATCA 960
Db 901 GCTAGTGAACCAAGAGCCCTTGACCTTGCCCGCTGAGCTATCCAGAAAGTGTCTATCA 960
QY 961 AGCGGCTACTAGACAGCCTGGGCGGCGGAGGCGGAGTTTGAGGGCGGCGTGTGAA 1020
Db 961 AGCGGCTACTAGACAGCCTGGGCGGCGGAGGCGGAGTTTGAGGGCGGCGTGTGAA 1020
QY 1021 TGTGTGCTGATCCAGACAAATGAGCAGTGGGTCTGTCTCCCTGAAGTCCCTGTGCC 1080
Db 1021 TGTGTGCTGATCCAGACAAATGAGCAGTGGGTCTGTCTCCCTGAAGTCCCTGTGCC 1080
QY 1081 CCGGGGTCCAGTCCCTCTTACATTTCTATCCCAAGGCCCGCTTCAAGTGTCCAGGG 1140
Db 1081 CCGGGGTCCAGTCCCTCTTACATTTCTATCCCAAGGCCCGCTTCAAGTGTCCAGGG 1140
QY 1141 AAGTGAAGGCGCTCTACAGTGTGAGCTTCTCTTGGGCGGAGATTCAGTGGGCC 1200
Db 1141 AAGTGAAGGCGCTCTCTACAGTGTGAGCTTCTCTTGGGCGGAGATTCAGTGGGCC 1200
QY 1201 AAGGATCTTTCAGCTTCTCCAGAGAGAAAGACTAGTTATGAGATTAAGCTCCAGA 1260
Db 1201 AAGGATCTTTCAGCTTCTCCAGAGAGAAAGACTAGTTATGAGATTAAGCTCCAGA 1260
QY 1261 GGCTTGGCCATCTATGTTGGGAAATTTCCCAAGACAAATGACAGCCCTATGTGATGG 1320
Db 1261 GGCTTGGCCATCTATGTTGGGAAATTTCCCAAGACAAATGACAGCCCTATGTGATGG 1320
QY 1321 AGGCTTTGGCATGGGCAAGTACACACGCGCTGACCGGTGGGGGAGACTGCGCTTACT 1380
Db 1321 AGGCTTTGGCATGGGCAAGTACACACGCGCTGACCGGTGGGGGAGACTGCGCTTACT 1380
QY 1381 GGCCACTTACGTGAGTGGCACTTCTTTTGGAGTCCAGAGGCCCAAGACAAATACGTGA 1440
Db 1381 GGCCACTTACGTGAGTGGCACTTCTTTTGGAGTCCAGAGGCCCAAGACAAATACGTGA 1440
QY 1441 TGCCCTTTTGTGTTTGAACAGAAACAGAGGCGCTCCCTGGCGGCAACCACTCAGATCT 1500
Db 1441 TGCCCTTTTGTGTTTGAACAGAAACAGAGGCGCTCCCTGGCGGCAACCACTCAGATCT 1500
QY 1501 CTACTGCGACTTCTTGGGGGCTTGGGAAAGGCTGTGCTGCTCAATCTATCTCAC 1560
Db 1501 CTACTGCGACTTCTTGGGGGCTTGGGAAAGGCTGTGCTGCTCAATCTATCTCAC 1560
QY 1561 CTGCTCAACTATGACTATGTGTGGATACGGTCTTCCACCCAGTGGGGCCATAGAAAT 1620
Db 1561 CTGCTCAACTATGACTATGTGTGGATACGGTCTTCCACCCAGTGGGGCCATAGAAAT 1620
QY 1621 ACGATTCTATGCGACGGGCTTACATAGCTCGGCAATTCCTTTGGTGTACTGGGAAATA 1680
Db 1621 ACGATTCTATGCGACGGGCTTACATAGCTCGGCAATTCCTTTGGTGTACTGGGAAATA 1680
QY 1681 CGGGAACCAAGTGTAGAGACACACCTGGGCAAGGTTCCACACCCAGCGCCACTTAA 1740
Db 1681 CGGGAACCAAGTGTAGAGACACACCTGGGCAAGGTTCCACACCCAGCGCCACTTAA 1740
QY 1741 GGTGATCTGATGTAGAGAGACTGAGAACTGGGCTTGGGCGGAGATATGCTTTGT 1800
Db 1741 GGTGATCTGATGTAGAGAGACTGAGAACTGGGCTTGGGCGGAGATATGCTTTGT 1800
QY 1801 CCCCATGGCTGTGCGCTTGAAGCCCTGAGACACCAAGTGTGAGAGGCTGACAGTACCAGG 1860
Db 1801 CCCCATGGCTGTGCGCTTGAAGCCCTGAGACACCAAGTGTGAGAGGCTGACAGTACCAGG 1860
QY 1861 GGTGCTGAGATGAGAGAGAGAGCGGCTTCTCTGTTGGGAAAGCGCCAGCCCTGCTACT 1920
Db 1861 GGTGCTGAGATGAGAGAGAGAGCGGCTTCTCTGTTGGGAAAGCGCCAGCCCTGCTACT 1920
QY 1921 GTACCTGCGCAGCAACACAGCAACAACTGGGGTACCCCGGGGCTACCGCATCAGAT 1980
Db 1921 GTACCTGCGCAGCAACACAGCAACAACTGGGGTACCCCGGGGCTACCGCATCAGAT 1980

OY 1981 GCTCAGCTTGTGCGAGAGCCCGCTGCCCCAAAACAGCTCCATGCGAGAGGCTTACGCTG 2040
|||||
Db 1981 GCTCAGCTTGTGCGAGAGCCCGCTGCCCCAAAACAGCTCCATGCGAGAGGCTTACGCTG 2040
OY 2041 GGAGAGGTACCAAGCTGGCGGTGACCCAGGGAAGGAGGAGGCCAGTAGCAGCAGCGCT 2100
|||||
Db 2041 GGAGAGGTACCAAGCTGGCGGTGACCCAGGGAAGGAGGAGGCCAGTAGCAGCAGCGCT 2100
OY 2101 TTTCATCATGAGATGACCTTTGGGCCCACTGAGATTTCAGTACGATTCATCAACATGA 2160
|||||
Db 2101 TTTCATCATGAGATGACCTTTGGGCCCACTGAGATTTCAGTACGATTCATCAACATGA 2160
OY 2161 GACCAATTGCTGGAAGAGATTTGGTGCCCTGGGTGACAGCTGGTTTTCTGCATATCCACGA 2220
|||||
Db 2161 GACCAATTGCTGGAAGAGATTTGGTGCCCTGGGTGACAGCTGGTTTTCTGCATATCCACGA 2220
OY 2221 TGCAGAGGACATTTCTTAACAGATGACTGTGGGGAACGGCGTGGCTTCTCTCCGAGC 2280
|||||
Db 2221 TGCAGAGGACATTTCTTAACAGATGACTGTGGGGAACGGCGTGGCTTCTCTCCGAGC 2280
OY 2281 CTATAACTCTTTGAGAGAACCCCTCTTCTACTCTGCGACTCCATCTACTTCCGAGG 2340
|||||
Db 2281 CTATAACTCTTTGAGAGAACCCCTCTTCTACTCTGCGACTCCATCTACTTCCGAGG 2340
OY 2341 GGACACGAGATGCTGGGGCTGCGAGGTCAACCCCTAGCTTGCTTCCGCCAGGCTGCTG 2400
|||||
Db 2341 GGACACGAGATGCTGGGGCTGCGAGGTCAACCCCTAGCTTGCTTCCGCCAGGCTGCTG 2400
OY 2401 CTGTGCCCCGACCTCCCTGCTCTTCCACAGGGGGCTTCTCCAACTAAGGCGGTCT 2460
|||||
Db 2401 CTGTGCCCCGACCTCCCTGCTCTTCCACAGGGGGCTTCTCCAACTAAGGCGGTCT 2460
OY 2461 GGGATGGGCGATGTGGCCCAAGGGCTCCAGGGCCAGGGGTGTGAGGAGTGGGGAGCAGCTG 2520
|||||
Db 2461 GGGATGGGCGATGTGGCCCAAGGGCTCCAGGGCCAGGGGTGTGAGGAGTGGGGAGCAGCTG 2520
OY 2521 GCACGTGGGCGGAGGCTGGTTCCTTCTTCTGTCGACGAGACTCTCTTCTTCCACTAC 2580
|||||
Db 2521 GCACGTGGGCGGAGGCTGGTTCCTTCTTCTGTCGACGAGACTCTCTTCTTCCACTAC 2580
OY 2581 CCTCCCTGCATCCGCTCTGAGCCAGAGGCTCTGACCTTGTGATGCTGTGACACAGG 2640
|||||
Db 2581 CCTCCCTGCATCCGCTCTGAGCCAGAGGCTCTGACCTTGTGATGCTGTGACACAGG 2640
OY 2641 GACACTGAACCTTGTGATGCCAGCTGTACTGAGTTCTCATCCACAGAGGCCAGGACATG 2700
|||||
Db 2641 GACACTGAACCTTGTGATGCCAGCTGTACTGAGTTCTCATCCACAGAGGCCAGGACATG 2700
OY 2701 CCCAGGCTGAGAGCGGTGGCCGAGGGCTTCCCTAGATGGTTCCCTTGTCTGTGCTG 2760
|||||
Db 2701 CCCAGGCTGAGAGCGGTGGCCGAGGGCTTCCCTAGATGGTTCCCTTGTCTGTGCTG 2760
OY 2761 TTCCCAATCTTTTAAAGGCACCTCCAAAGGACTTAAGAGGGGCTTTCCCTGGAGAC 2820
|||||
Db 2761 TTCCCAATCTTTTAAAGGCACCTCCAAAGGACTTAAGAGGGGCTTTCCCTGGAGAC 2820
OY 2821 CCAGAGTAGGGTTCAGAGTCTGCAGATGCCATAGCTGAGTGAAGAGATGCTTCTGCTC 2880
|||||
Db 2821 CCAGAGTAGGGTTCAGAGTCTGCAGATGCCATAGCTGAGTGAAGAGATGCTTCTGCTC 2880
OY 2881 ACATTTCCCTCTCATCCAGGCTTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2940
|||||
Db 2881 ACATTTCCCTCTCATCCAGGCTTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2940
OY 2941 TCT 3000
|||||
Db 2941 TCT 3000
OY 3001 ATCCCTATGTGCCAGCCCTGTGACTCCCGAGCCCTCAATTTTCACTCAAGTTCCGCTCT 3060
|||||
Db 3001 ATCCCTATGTGCCAGCCCTGTGACTCCCGAGCCCTCAATTTTCACTCAAGTTCCGCTCT 3060
OY 3061 CCTCTCAGCCCTATGGAATCTCAAGTCAAGGACGAGCCCTAATCAAGTGGCCAAATCC 3120

|||||
Db 3061 CCTCTCAGCCCTATGGAATCTCAAGTCAAGGACGAGCCCTAATCAAGATGGCCAAATCCC 3120
OY 3121 TGTGTGCTGCTTCCCTTGTGCTGTGCTTATTTGGAGTAGAGTGTCTCTACCCGCTGTC 3180
|||||
Db 3121 TGTGTGCTGCTTCCCTTGTGCTGTGCTTATTTGGAGTAGAGATTTGCTCTACCCGCTGTC 3180
OY 3181 CTGGGCTGGGTGTCTTTCAGAGACGCTGCTTCTGTGCAATTTGTCTGTGCTCTCTGATG 3240
|||||
Db 3181 CTGGGCTGGGTGTCTTTCAGAGACGCTGCTTCTGTGCAATTTGTCTGTGCTCTCTGATG 3240
OY 3241 CTCTCTATAGAGAGAGATGCTCATCTGTGACAGACGACTCAAGTTACATTTCAGTGA 3300
|||||
Db 3241 CTCTCTATAGAGAGAGATGCTCATCTGTGACAGACGACTCAAGTTACATTTCAGTGA 3300
OY 3301 TTTGGGGGTGCATGATTAATGAAGAAATGGCCATTTTGTACAGAGGCTCTGATTTCTGCA 3360
|||||
Db 3301 TTTGGGGGTGCATGATTAATGAAGAAATGGCCATTTTGTACAGAGGCTCTGATTTCTGCA 3360
OY 3361 CAGCCTGTTTGGAGGCTGAGTGAAGAAACAAAGGCTGGCATCAAGATGAGAACCCAAA 3420
|||||
Db 3361 CAGCCTGTTTGGAGGCTGAGTGAAGAAACAAAGGCTGGCATCAAGATGAGAACCCAAA 3420
OY 3421 GCCCTCTCAACTCAGACCAACCCAGCAGAGAGGGGCTGTCCAAATCACTTACAGCATGCGA 3480
|||||
Db 3421 GCCCTCTCAACTCAGACCAACCCAGCAGAGAGGGGCTGTCCAAATCACTTACAGCATGCGA 3480
OY 3481 ATGAGCTGGGCTCTGGGTGAGTGGGGGTCTGGGCTTGTGAGGAGGGGCTGGGCTGGGT 3540
|||||
Db 3481 ATGAGCTGGGCTCTGGGTGAGTGGGGGTCTGGGCTTGTGAGGAGGGGCTGGGCTGGGT 3540
OY 3541 GGGGACAGGGCTGGCTGTGTCAGAGGCTTGGGCTCCATTTCCATCATCTGTGCTGCTCTG 3600
|||||
Db 3541 GGGGACAGGGCTGGCTGTGTCAGAGGCTTGGGCTCCATTTCCATCATCTGTGCTGCTCTG 3600
OY 3601 AGGCTGATTTGGGGATGGGGAGCAAGAAATPAGCAAGATGAGAAACCAAGAACTTT 3660
|||||
Db 3601 AGGCTGATTTGGGGATGGGGAGCAAGAAATPAGCAAGATGAGAAACCAAGAACTTT 3660
OY 3661 TTTCTCTAAAGCATGCTGTTAAATCAATTTCTGATACAGCTTACAAATAGATGACAG 3720
|||||
Db 3661 TTTCTCTAAAGCATGCTGTTAAATCAATTTCTGATACAGCTTACAAATAGATGACAG 3720
OY 3721 CTAAAAAATATGTATGCTTTATATATCTAATATGTAATCTTACAGTGAAGAAAGGC 3780
|||||
Db 3721 CTAAAAAATATGTATGCTTTATATATCTAATATGTAATCTTACAGTGAAGAAAGGC 3780
OY 3781 AAGCCACAGAAATGTATAGCCGACCTCCCATTTGTTTCCAGAAAGAGTGAATATA 3840
|||||
Db 3781 AAGCCACAGAAATGTATAGCCGACCTCCCATTTGTTTCCAGAAAGAGTGAATATA 3840
OY 3841 AACACATATTTGCTTATGATGCTTATTCAGATTAATAGGTAACACTGATTTATTTGG 3900
|||||
Db 3841 AACACATATTTGCTTATGATGCTTATTCAGATTAATAGGTAACACTGATTTATTTGG 3900
OY 3901 GAGGGGAACCAAGTAGTTGAGAGACAGAGAGAGAGGGTCTTAACTTACACCTTTTG 3960
|||||
Db 3901 GAGGGGAACCAAGTAGTTGAGAGACAGAGAGAGAGGGTCTTAACTTACACCTTTTG 3960
OY 3961 TACATTTGAATTTGAACCATGTGACTGATTAATCAATTAATTAATTAATTAATGAGG 4020
|||||
Db 3961 TACATTTGAATTTGAACCATGTGACTGATTAATCAATTAATTAATTAATTAATGAGG 4020
OY 4021 CCCAAAAAAAAAAAAAAAAAAAA 4040
|||||
Db 4021 CCCAAAAAAAAAAAAAAAAAAAA 4040

RESULT 2
AX333039 4040 bp DNA linear PAT 09-JAN-2002
LOCUS
DEFINITION Sequence 3548 from Patent W00194629.
ACCESSION AX333039

VERSION AX33039.1 GI:18123673
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Young, P. E., Augustus, M., Carter, K. C., Ebner, R., Endress, G.,
Horligan, S., Sopper, D. R. and Weaver, Z.
Cancer gene determination and therapeutic screening using signature
gene sets
Patent: WO 0194629-A 3548 13-DEC-2001;
JOURNAL
Avalon Pharmaceuticals (US)
FEATURES
source 1..4040
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 825 a 1183 c 1097 g 934 t 1 others
ORIGIN
Query Match 100.0%; Score 4039.6; DB 6; Length 4040;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4039; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
1 GTCCTCCACACCTTATGTCAGGATCTGACTACCGGGAACCTCAGCAGACCGGGA 60
1 GTCTTCCACACCTTATGTCAGGATCTGACTACCGGGAACCTCAGCAGACCGGGA 60
61 GCCCCCCACCCGTCAGAGACCAAGAGCCCGCTTCTGCTGCGTGAATATCATTTG 120
61 GCCCCCCACCCGTCAGAGACCAAGAGCCCGCTTCTGCTGCGTGAATATCATTTG 120
121 CTCTCTTTGGTGAATCAGTGTCTCTTCTGCGGAAAATGAACAGAGAACATCT 180
121 CTCTCTTTGGTGAATCAGTGTCTCTTCTGCGGAAAATGAACAGAGAACATCT 180
181 CGTCTCTCTATCTGCGCGCTATCAGCATCTTGCCTTGTGTGTGTCCTGCTGGTGG 240
181 CGTCTCTCTATCTGCGCGCTATCAGCATCTTGCCTTGTGTGTGTCCTGCTGGTGG 240
181 CGTCTCTCTATCTGCGCGCTATCAGCATCTTGCCTTGTGTGTGTCCTGCTGGTGG 240
241 CAGGGGTGAGATGGGGTGAACCCAGCAGCTTCCCATTTGCCCTCTGTATCTCCAG 300
241 CAGGGGTGAGATGGGGTGAACCCAGCAGCTTCCCATTTGCCCTCTGTATCTCCAG 300
241 CAGGGGTGAGATGGGGTGAACCCAGCAGCTTCCCATTTGCCCTCTGTATCTCCAG 300
301 TGCCCCAGCTTGGACACACCTGGCCAGAGCAGCTGTTTGCAGACCTGAGCCAGAGA 360
301 TGCCCCAGCTTGGACACACCTGGCCAGAGCAGCTGTTTGCAGACCTGAGCCAGAGA 360
361 GCTACGAGCTGATGAGCTTTCGACCCAGCGCTGGGGCCAGGGCTGTGTGATGACG 420
361 GCTACGAGCTGATGAGCTTTCGACCCAGCGCTGGGGCCAGGGCTGTGTGATGACG 420
421 CCAGGCCCCGCTCGAGCAACTGTCTTCTCAGTGAAGTTCAGCTGCCCTCCCAAGG 480
421 CCAGGCCCCGCTCGAGCAACTGTCTTCTCAGTGAAGTTCAGCTGCCCTCCCAAGG 480
421 CCAGGCCCCGCTCGAGCAACTGTCTTCTCAGTGAAGTTCAGCTGCCCTCCCAAGG 480
421 CCAGGCCCCGCTCGAGCAACTGTCTTCTCAGTGAAGTTCAGCTGCCCTCCCAAGG 480
481 TGCAAGCTTGGCTACCTTGGACAGAGGGAGAGCCCTGCTGCGGAGAGCATGGCCAT 540
481 TGCAAGCTTGGCTACCTTGGACAGAGGGAGAGCCCTGCTGCGGAGAGCATGGCCAT 540
481 TGCAAGCTTGGCTACCTTGGACAGAGGGAGAGCCCTGCTGCGGAGAGCATGGCCAT 540
481 TGCAAGCTTGGCTACCTTGGACAGAGGGAGAGCCCTGCTGCGGAGAGCATGGCCAT 540
541 CGTCTTCTTGGACAGCAACCCAGCAGCTGAGTGAAGTGGGGGCGCCACTGCC 600
541 CGTCTTCTTGGACAGCAACCCAGCAGCTGAGTGAAGTGGGGGCGCCACTGCC 600
541 CGTCTTCTTGGACAGCAACCCAGCAGCTGAGTGAAGTGGGGGCGCCACTGCC 600
601 TCACCCCTCTACATGCGGAGCTGATGTGAGAGCTCATGAGAGCCCTGCGCTATCA 660
601 TCACCCCTCTACATGCGGAGCTGATGTGAGAGCTCATGAGAGCCCTGCGCTATCA 660
601 TCACCCCTCTACATGCGGAGCTGATGTGAGAGCTCATGAGAGCCCTGCGCTATCA 660
661 CCGAGCCCGCTGCTTTCAGAGTACCTGAGATAGACAGATGATCTTACAGAGA 720
661 CCGAGCCCGCTGCTTTCAGAGTACCTGAGATAGACAGATGATCTTACAGAGA 720
661 CCGAGCCCGCTGCTTTCAGAGTACCTGAGATAGACAGATGATCTTACAGAGA 720
721 GCTGCCCAAGGCTTGTGGGCTTCTCAGCAGCTGTTGCTTACAGACCGGGAGGAA 780

|||||
Db 721 GCTGCCCAAGGCTTGTGGGCTTCTCAGCAGCTGTTGCTTACAGACCGGGAGGAA 780
|||||
Qy 781 CCGTGTGACATGACACAGGCTCCCGTGTCTGCAATCAGGGAGCCGGCCACTGCTT 840
781 CCGTGTGACATGACACAGGCTCCCGTGTCTGCAATCAGGGAGCCGGCCACTGCTT 840
841 TGGGCTTACTACAACTTCGCGGCTGCTTCTTCTCAGCAGCTGCTGCTGAGCT 900
841 TGGGCTTACTACAACTTCGCGGCTGCTTCTTCTCAGCAGCTGCTGCTGAGCT 900
841 TGGGCTTACTACAACTTCGCGGCTGCTTCTTCTCAGCAGCTGCTGCTGAGCT 900
901 GCTAGTGAACACAGGCTTGAACCTGAGCTGAGCTATCCAGAGAGTGTCTATCA 960
901 GCTAGTGAACACAGGCTTGAACCTGAGCTGAGCTATCCAGAGAGTGTCTATCA 960
901 GCTAGTGAACACAGGCTTGAACCTGAGCTGAGCTATCCAGAGAGTGTCTATCA 960
961 AGGCGCTACTACAGACAGCTGCGCCAGCTGAGAGCCCAAGTTGAGGCGCGCTGTGAA 1020
961 AGGCGCTACTACAGACAGCTGCGCCAGCTGAGAGCCCAAGTTGAGGCGCGCTGTGAA 1020
1021 TGTGCTCTATCCAGACAAATGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1080
1021 TGTGCTCTATCCAGACAAATGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1080
1081 CCGGCTCAGCTCCCGCTCTACAGTCTATCCCAAGAGCCCGCTTCAAGTGTGAGG 1140
1081 CCGGCTCAGCTCCCGCTCTACAGTCTATCCCAAGAGCCCGCTTCAAGTGTGAGG 1140
1141 AAGTGAAGTGGCTCTCTACTGTGAGCTTCTCTTGGCTGAGAGATTCAGTGGCCC 1200
1141 AAGTGAAGTGGCTCTCTACTGTGAGCTTCTCTTGGCTGAGAGATTCAGTGGCCC 1200
1201 AAGGATCTTGAAGTGTGGCTTCCCAAGAGAGAAAGCTAGTTATAGATTAAGCTTCAAGA 1260
1201 AAGGATCTTGAAGTGTGGCTTCCCAAGAGAGAAAGCTAGTTATAGATTAAGCTTCAAGA 1260
1201 AAGGATCTTGAAGTGTGGCTTCCCAAGAGAGAAAGCTAGTTATAGATTAAGCTTCAAGA 1260
1261 GGCCTTGGCTATGAGTGAAGTATCCCGAGAGATGAGAGAGAGAGAGAGAGAGAGAG 1320
1261 GGCCTTGGCTATGAGTGAAGTATCCCGAGAGATGAGAGAGAGAGAGAGAGAGAGAG 1320
1321 AGGCTTGGCTATGAGTGAAGTATCCCGAGAGATGAGAGAGAGAGAGAGAGAGAGAG 1380
1321 AGGCTTGGCTATGAGTGAAGTATCCCGAGAGATGAGAGAGAGAGAGAGAGAGAGAG 1380
1381 GGCACACAGTGAAGTGAAGTATCCCGAGAGATGAGAGAGAGAGAGAGAGAGAGAG 1440
1381 GGCACACAGTGAAGTGAAGTATCCCGAGAGATGAGAGAGAGAGAGAGAGAGAGAG 1440
1441 TGCCCTTGTGTGTGAAG 1500
1441 TGCCCTTGTGTGTGAAG 1500
1501 CTACTGCACTACTTGGGGCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
1501 CTACTGCACTACTTGGGGCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
1561 CTGCTCACTACTTGGGGCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
1561 CTGCTCACTACTTGGGGCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
1621 AGCATTCATATGAG 1680
1621 AGCATTCATATGAG 1680
1681 CCGGAGACCAAGTGTGAG 1740
1681 CCGGAGACCAAGTGTGAG 1740
1741 GGTGATATGATATGAG 1800
1741 GGTGATATGATATGAG 1800
1801 CCGCATGCTGCTGCTGAG 1860
|||||

```
Db 1801 CCCCATGGCTGTGTCCTGGAGCCCTGAGCACACAGCTGCAGAGGCTGCAGGCTGACCCGGAA 1860
Qy 1861 GCTGTGAGAGATGAGAGAGAGCCGCTTCTCTGTGGAAAGCGCACCCCTGCTACCT 1920
Db 1861 GCTGTGAGAGATGAGAGAGAGCCGCTTCTCTGTGGAAAGCGCACCCCTGCTACCT 1920
Qy 1921 GTACCTGGCCAGCAACACAGCAACAAGTGGGGTCAACCCCGGGGGCTACCCGATCCAGAT 1980
Db 1921 GTACCTGGCCAGCAACACAGCAACAAGTGGGGTCAACCCCGGGGGCTACCCGATCCAGAT 1980
Qy 1981 GTCACAGCTTGTGAGAGAGCCGCTGCCCCAAAACAGCTCCATGGCCGAGAGCTTCAGCTG 2040
Db 1981 GTCACAGCTTGTGAGAGAGCCGCTGCCCCAAAACAGCTCCATGGCCGAGAGCTTCAGCTG 2040
Qy 2041 GGAGAGGTACACAGCTGTGCTGTGACCCAGCCGAGAGAGAGAGAGCCAGTAGACAGACAGCT 2100
Db 2041 GGAGAGGTACACAGCTGTGCTGTGACCCAGCCGAGAGAGAGAGAGCCAGTAGACAGACAGCT 2100
Qy 2101 TTTCAATCAGAAATGACCTTGGGGCCCCACTGTGGATTGAGTACTTCAATCAACAAATGA 2160
Db 2101 TTTCAATCAGAAATGACCTTGGGGCCCCACTGTGGATTGAGTACTTCAATCAACAAATGA 2160
Qy 2161 GACCATGTGTGGAAGAGATTGGTGGCTGGGTGACAGCTGTGTTTGTGATATCCACA 2220
Db 2161 GACCATGTGTGGAAGAGATTGGTGGCTGGGTGACAGCTGTGTTTGTGATATCCACA 2220
Qy 2221 TGCAGAGGACATTCCTAACACAGTACTGTGGGGAACGGCGTGGCTTCTCTCGACCC 2280
Db 2221 TGCAGAGGACATTCCTAACACAGTACTGTGGGGAACGGCGTGGCTTCTCTCGACCC 2280
Qy 2281 CTATTAATCTTTGTGAGCAAGACCCCTCTTACTCTGCGAGCACTCAATCACTTCGAGG 2340
Db 2281 CTATTAATCTTTGTGAGCAAGACCCCTCTTACTCTGCGAGCACTCAATCACTTCGAGG 2340
Qy 2341 GGACACAGAGTCTGGGGCTGTGCGAGGTCAACCCCTAGCTTGGCTGCCAGCTGTGTC 2400
Db 2341 GGACACAGAGTCTGGGGCTGTGCGAGGTCAACCCCTAGCTTGGCTGCCAGCTGTGTC 2400
Qy 2401 CTGTGCCCCGACCTCCCTGCTTCTCCACGGGGCTTCTCTACAACTAGGGGCTCT 2460
Db 2401 CTGTGCCCCGACCTCCCTGCTTCTCCACGGGGCTTCTCTACAACTAGGGGCTCT 2460
Qy 2461 GGGATGGGGGATGTGGGCAAGGGGCTCCAGGGCCAGGGGTGAGAGGAGGGAGAGAGCTGG 2520
Db 2461 GGGATGGGGGATGTGGGCAAGGGGCTCCAGGGCCAGGGGTGAGAGGAGGGAGAGAGCTGG 2520
Qy 2521 GCACTGGGGCCGACAGCTGTGTTCCCTTCTGTGCGAGAGCTCTTTTCTCCACTAC 2580
Db 2521 GCACTGGGGCCGACAGCTGTGTTCCCTTCTGTGCGAGAGCTCTTTTCTCCACTAC 2580
Qy 2581 CCTCCCTGCGATCCGCTCTAGCCAGAGAGCTCTGACCTGTGATCCCTGACACAGGG 2640
Db 2581 CCTCCCTGCGATCCGCTCTAGCCAGAGAGCTCTGACCTGTGATCCCTGACACAGGG 2640
Qy 2641 GACACTGAACCTTTGATGCGACGCTGACTGAGTTCATCAACAGAGGGCCAGGCAATGG 2700
Db 2641 GACACTGAACCTTTGATGCGACGCTGACTGAGTTCATCAACAGAGGGCCAGGCAATGG 2700
Qy 2701 CCCAGCTGTAGAGCCGTGGCCAGGGCTTCCCTAGATGGTTCCTTTGTGTGCTGTGCT 2760
Db 2701 CCCAGCTGTAGAGCCGTGGCCAGGGCTTCCCTAGATGGTTCCTTTGTGTGCTGTGCT 2760
Qy 2761 TTCCCGAATCTTTTAAAGCCACTCCAAAGACTTAAAGGGGGCTATTCCCTGGAGACC 2820
Db 2761 TTCCCGAATCTTTTAAAGCCACTCCAAAGACTTAAAGGGGGCTATTCCCTGGAGACC 2820
Qy 2821 CCAAGATAGGCTTGCAGATCTCTGCAAGTCACTAGCTGAGCTGGAAGAGATGCTTCTGCTC 2880
Db 2821 CCAAGATAGGCTTGCAGATCTCTGCAAGTCACTAGCTGAGCTGGAAGAGATGCTTCTGCTC 2880
Qy 2881 ACATTTCCCTCATCCAGAGCTTCTCTCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2940
Db 2881 ACATTTCCCTCATCCAGAGCTTCTCTCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2940
Qy 2941 TCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3000
Db 2941 TCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3000
Qy 3001 ATGCCATATGTCCAGAGCCCTGTACTCTCCAGAGCCCTCACTTTTCAAGTTCGCTCT 3060
Db 3001 ATGCCATATGTCCAGAGCCCTGTACTCTCCAGAGCCCTCACTTTTCAAGTTCGCTCT 3060
Qy 3061 CCTCTCCAGCCCTATGAGAGTCTCAAGTCAAGGAGCCCTTAATCAGAGTGGCCAAATCCC 3120
Db 3061 CCTCTCCAGCCCTATGAGAGTCTCAAGTCAAGGAGCCCTTAATCAGAGTGGCCAAATCCC 3120
Qy 3121 TGTGTGTGCTTCCCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3180
Db 3121 TGTGTGTGCTTCCCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3180
Qy 3181 CTGGGGCTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3240
Db 3181 CTGGGGCTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3240
Qy 3241 CTCTCTATAGAGAGATGTGTATCTGTACAGCAGCAGCTCAAGTATAGCATTTTCAAGTGA 3300
Db 3241 CTCTCTATAGAGAGATGTGTATCTGTACAGCAGCAGCTCAAGTATAGCATTTTCAAGTGA 3300
Qy 3301 TTTGGGGGTGCAATGATTAATGAAGATGGCCATTTTGTACAGAGGCTGTATTTCTGCAA 3360
Db 3301 TTTGGGGGTGCAATGATTAATGAAGATGGCCATTTTGTACAGAGGCTGTATTTCTGCAA 3360
Qy 3361 CAGCCTGTTTGGAGCTGTGAGTGTGAAGAAAGGGTGGGCAATCAAGATGAAGAACCAAA 3420
Db 3361 CAGCCTGTTTGGAGCTGTGAGTGTGAAGAAAGGGTGGGCAATCAAGATGAAGAACCAAA 3420
Qy 3421 GCCCCCTACACTCCAGCACCACCCAGCAGAGAGGGCTGTCCATCATTACAGCATTCAGCATGCA 3480
Db 3421 GCCCCCTACACTCCAGCACCACCCAGCAGAGAGGGCTGTCCATCATTACAGCATTCAGCATGCA 3480
Qy 3481 ATGAGCTGTGGGCTGTGGGTGAGTGGGGGTGTGGGCTGTGGGAGGGGGCTGTGGGCTGTGG 3540
Db 3481 ATGAGCTGTGGGCTGTGGGTGAGTGGGGGTGTGGGCTGTGGGAGGGGGCTGTGGGCTGTGG 3540
Qy 3541 GGGGAGGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGG 3600
Db 3541 GGGGAGGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGG 3600
Qy 3601 AGCTCTGAGTGTGGGATGGGGAACAAGAAATAGCAAGATGAGAGAAACAGAAACTTT 3660
Db 3601 AGCTCTGAGTGTGGGATGGGGAACAAGAAATAGCAAGATGAGAGAAACAGAAACTTT 3660
Qy 3661 TTTCTCTAAGAGATGCTGTTAATCAATCTCTATACAGCTTCAATCAATCAATCAATCAATCA 3720
Db 3661 TTTCTCTAAGAGATGCTGTTAATCAATCTCTATACAGCTTCAATCAATCAATCAATCAATCA 3720
Qy 3721 CTAAAAAATTAATGTATGCTCTTATATATATATATATATATATATATATATATATATATATAT 3780
Db 3721 CTAAAAAATTAATGTATGCTCTTATATATATATATATATATATATATATATATATATATATAT 3780
Qy 3781 AAGCCACAGAAATGTATATAGCGCACTTCCCAATTTGTGTTTCAGAAAGAGATAGAAATYA 3840
Db 3781 AAGCCACAGAAATGTATATAGCGCACTTCCCAATTTGTGTTTCAGAAAGAGATAGAAATYA 3840
Qy 3841 AACACATATATGCTTATATATATATATATATATATATATATATATATATATATATATATAT 3900
Db 3841 AACACATATATGCTTATATATATATATATATATATATATATATATATATATATATATATAT 3900
Qy 3901 GAGGGGAACAGTATGTTGAGAGACAGGAGAGAGAGGGTCTTAACATTAACCTTTG 3960
Db 3901 GAGGGGAACAGTATGTTGAGAGACAGGAGAGAGAGGGTCTTAACATTAACCTTTG 3960
Qy 3961 TACATTTTGAATTTTGAACCATGTGACTGTATTAACCTTTTGAATTAATTAATTAATTAATTA 4020
Db 3961 TACATTTTGAATTTTGAACCATGTGACTGTATTAACCTTTTGAATTAATTAATTAATTAATTA 4020
```

```
QY 4021 CCCAAAAAAAAAAAAAAAAA 4040
    |||||
Db 4021 CCCAAAAAAAAAAAAAAAAA 4040
```

RESULT 3	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
HSU39447	4040 bp	Human placenta copper monamine oxidase mRNA, complete cds.	U39447	GI:1399031		Homo sapiens.	Homo sapiens

CDS

sig_peptide	161.	.217	/note="possible signal peptide"
misc_feature	1571.	.1573	/note="encodes Tyr residue that is converted to the topa
			quinone cofactor"
polyA_signal	4004.	.4009	
polyA_signal	4011.	.4016	
BASE COUNT	825	a 1183 c 1097 g 934 t 1	others
ORIGIN			

Query Match	100.0%	Score 4039.6	DB 9	Length 4040
Best Local Similarity	100.0%	Pred. NO. 0		
Matches 4039: Conservative	1	Mismatches	0	Indels 0; Gaps 0;
0y	1	GTCTTCCACCCCTTAGTCCAGCAGCATCTGATACCGGGAACCTTAGCCAGAGTCCGGGA	60	

[illegible]

QY 1141 AAGTGAAGTGGCTCTCTCACTGTGACTTTCTCTTGGCCCTGGAGCATTCAGTGGCC 1200
| | | | |
Db 1141 AAGTGAAGTGGCTCTCTCACTGTGACTTTCTCTTGGCCCTGGAGCATTCAGTGGCC 1200
QY 1201 AAGGATCTTGAAGTGGCTCTCAAGAGAGAAAGACTAGTTATGATATAAGCTCCAGA 1260
| | | | |
Db 1201 AAGGATCTTGAAGTGGCTCTCAAGAGAGAAAGACTAGTTATGATATAAGCTCCAGA 1260
QY 1261 GGCCTGGCCATCTATGTGGAATTTCCAGAGCAATGACAGCCGCTATGTGATG 1320
| | | | |
Db 1261 GGCCTGGCCATCTATGTGGAATTTCCAGAGCAATGACAGCCGCTATGTGATG 1320
QY 1321 AAGCTTTGGCATGGGCAAGTACACACGCCCCGTGACCCGTGGGTGAGTGCCTTACTT 1380
| | | | |
Db 1321 AAGCTTTGGCATGGGCAAGTACACACGCCCCGTGACCCGTGGGTGAGTGCCTTACTT 1380
QY 1381 GGCACCTACGTGAGTGGCACTTCTTTGAGTCCAGAGGCCCCCAAGACATTAAGTGA 1440
| | | | |
Db 1381 GGCACCTACGTGAGTGGCACTTCTTTGAGTCCAGAGGCCCCCAAGACATTAAGTGA 1440
QY 1441 TGCCTTTGTGTGTGGAACAGAACGAGGCTCCCTGCGGCAACACCTCAGATCT 1500
| | | | |
Db 1441 TGCCTTTGTGTGTGGAACAGAACGAGGCTCCCTGCGGCAACACCTCAGATCT 1500
QY 1501 CTACTGGCACTACTTTGGGGGCTTGGGGAAGGGTGGTGGTGCATCTATGTCAC 1560
| | | | |
Db 1501 CTACTGGCACTACTTTGGGGGCTTGGGGAAGGGTGGTGGTGCATCTATGTCAC 1560
QY 1561 CTGTCTCACTATGACTATGTGGGATACGCTTTCACACCCAGTGGGGCCATAGAAAT 1620
| | | | |
Db 1561 CTGTCTCACTATGACTATGTGGGATACGCTTTCACACCCAGTGGGGCCATAGAAAT 1620
QY 1621 ACGATTCTATGCCACGAGGCTACATCAGCTGGCATCTCTTTGGTCTACTGGGAAT 1680
| | | | |
Db 1621 ACGATTCTATGCCACGAGGCTACATCAGCTGGCATCTCTTTGGTCTACTGGGAAT 1680
QY 1681 CGGGAACCAAGTGTGAGAGACACACCCCTGGGACAGGTCACACCCAGCCCACTTCAA 1740
| | | | |
Db 1681 CGGGAACCAAGTGTGAGAGACACACCCCTGGGACAGGTCACACCCAGCCCACTTCAA 1740
QY 1741 GGTGATCTGTGATGTAGACAGGATGTGAACTGGGTCTGGGCGAGGATATGCTTGT 1800
| | | | |
Db 1741 GGTGATCTGTGATGTAGACAGGATGTGAACTGGGTCTGGGCGAGGATATGCTTGT 1800
QY 1801 CCCCATGGCTGTCCCTGTGAGCCCTGAGCACACAGCTGCAGAGCTGAGTGAACCCGAA 1860
| | | | |
Db 1801 CCCCATGGCTGTCCCTGTGAGCCCTGAGCACACAGCTGCAGAGCTGAGTGAACCCGAA 1860
QY 1861 GGTGCTGGAGATGAGAGAGAGGCGCCCTCTCTGTTGGGAAGCGCCACCCCTGCTACT 1920
| | | | |
Db 1861 GGTGCTGGAGATGAGAGAGAGGCGCCCTCTCTGTTGGGAAGCGCCACCCCTGCTACT 1920
QY 1921 GTACTGGCCAGCAACACAGCAAAAGTGGGTCAACCCCGGGGCTACCGCATCCAGAT 1980
| | | | |
Db 1921 GTACTGGCCAGCAACACAGCAAAAGTGGGTCAACCCCGGGGCTACCGCATCCAGAT 1980
QY 1981 GCTCAGCTTTGTGTGAGAGCCGCTGCCCAAAAGCTCCTCATGTGCGAGAGCTTCACTG 2040
| | | | |
Db 1981 GCTCAGCTTTGTGTGAGAGCCGCTGCCCAAAAGCTCCTCATGTGCGAGAGCTTCACTG 2040
QY 2041 GGAAGGTACAGCTGGCTGTGACCCAGGGAAGGAGAGGAGCCGATGACAGCAGCT 2100
| | | | |
Db 2041 GGAAGGTACAGCTGGCTGTGACCCAGGGAAGGAGAGGAGCCGATGACAGCAGCAGT 2100
QY 2101 TTTCATCAGATATACCTTTGGGCCCCACTGTGTGATTTAGTACTTCAATCAATGA 2160
| | | | |
Db 2101 TTTCATCAGATATACCTTTGGGCCCCACTGTGTGATTTAGTACTTCAATCAATGA 2160
QY 2161 GACCATTTGTGGAAGGATTTGTGGCTGGGTGACAGCTGTGTTTCTGCATATCCACA 2220
| | | | |
Db 2161 GACCATTTGTGGAAGGATTTGTGGCTGGGTGACAGCTGTGTTTCTGCATATCCACA 2220

QY 2221 TGCAGAGCATTTCTTACACAGTACTGAGGGAAGGGGGTGGGCTTCTCCAGCC 2280
| | | | |
Db 2221 TGCAGAGCATTTCTTACACAGTACTGAGGGAAGGGGGTGGGCTTCTCCAGCC 2280
QY 2281 CTATACTTCTTTGAGAGAGACCCCTCTTCTACTCTGCGCATCTACTTCTCCAGG 2340
| | | | |
Db 2281 CTATACTTCTTTGAGAGAGACCCCTCTTCTACTCTGCGCATCTACTTCTCCAGG 2340
QY 2341 GGAACAGATGTGGGGCTGTGAGGTCAACCCCTAGCTGCTGCCGCCAGCTCTG 2400
| | | | |
Db 2341 GGAACAGATGTGGGGCTGTGAGGTCAACCCCTAGCTGCTGCCGCCAGCTCTG 2400
QY 2401 CTGTGCCCCGACCTCCCTGCTTCTCCAGCGGGGGCTTCTCACAACATGAGCGGCT 2460
| | | | |
Db 2401 CTGTGCCCCGACCTCCCTGCTTCTCCAGCGGGGGCTTCTCACAACATGAGCGGCT 2460
QY 2461 GGGATGGGATGTGGCCAAAGGCTCCAGGGCCAGGGTGTGAGGATGGGAGACAGCTG 2520
| | | | |
Db 2461 GGGATGGGATGTGGCCAAAGGCTCCAGGGCCAGGGTGTGAGGATGGGAGACAGCTG 2520
QY 2521 GCACTGGGCGGAGGCTGTGCTTCTTCTGTCGAGAGACTCTTCTTCTTCCACTAC 2580
| | | | |
Db 2521 GCACTGGGCGGAGGCTGTGCTTCTTCTGTCGAGAGACTCTTCTTCTTCCACTAC 2580
QY 2581 CTCCCTCGCATCCGCTGTGAGCCAGAGACCTCTGACCTGATGATGCTGACACAGG 2640
| | | | |
Db 2581 CTCCCTCGCATCCGCTGTGAGCCAGAGACCTCTGACCTGATGATGCTGACACAGG 2640
QY 2641 GACACTGAACCTTGTGTATGCCAGCTGTACTGATGATTCATCCACAGAGGCCAGGATG 2700
| | | | |
Db 2641 GACACTGAACCTTGTGTATGCCAGCTGTACTGATGATTCATCCACAGAGGCCAGGATG 2700
QY 2701 CCCAGCTGAGAGCGGTGGGAGGCTTCCCTAGATGCTTCTTGTGTGCTGTGCT 2760
| | | | |
Db 2701 CCCAGCTGAGAGCGGTGGGAGGCTTCCCTAGATGCTTCTTGTGTGCTGTGCT 2760
QY 2761 TTCCCGAATCTTTTGAAGGACCTCCAGAGACTTAAAGGGGGCTTTCCTGAGACCC 2820
| | | | |
Db 2761 TTCCCGAATCTTTTGAAGGACCTCCAGAGACTTAAAGGGGGCTTTCCTGAGACCC 2820
QY 2821 CCAGAGTGGGTTGCCAGTCTCTGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 2880
| | | | |
Db 2821 CCAGAGTGGGTTGCCAGTCTCTGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 2880
QY 2881 ACATTCCTCTCATCCAGGCTCTTCTCTCTGCTTCTCTCTCTCTCTCTCTCTCTCT 2940
| | | | |
Db 2881 ACATTCCTCTCATCCAGGCTCTTCTCTCTGCTTCTCTCTCTCTCTCTCTCTCTCT 2940
QY 2941 TCCCT 3000
| | | | |
Db 2941 TCCCT 3000
QY 3001 ATCCCTATGTCCAGCCCTGTGACTCCCCAGCCCTCAAGTTTCACTCAAGTCCGCT 3060
| | | | |
Db 3001 ATCCCTATGTCCAGCCCTGTGACTCCCCAGCCCTCAAGTTTCACTCAAGTCCGCT 3060
QY 3061 CCTCTCCAGCCCTTATGAAAGTCTAAGGTACAGGAGACCCCTAATCAAGAGTGGCAATCC 3120
| | | | |
Db 3061 CCTCTCCAGCCCTTATGAAAGTCTAAGGTACAGGAGACCCCTAATCAAGAGTGGCAATCC 3120
QY 3121 TGT 3180
| | | | |
Db 3121 TGT 3180
QY 3181 CTGGGGCTGGGT 3240
| | | | |
Db 3181 CTGGGGCTGGGT 3240
QY 3241 CTCTCTTATAGAGAGATGT 3300
| | | | |
Db 3241 CTCTCTTATAGAGAGATGT 3300
QY 3301 TTGCGGGGTGCAATGATTAAGAAATGGCCATTTGTACAGGGGCTGTATTTCTGCAA 3360
| | | | |

Db	3301	TTTGGGGTGC	AAAGATTAATGAAG	ATGGCCATTTT	GTACAGGGCTCTGT	TTTCCAA	3360
QY	3361	CAGCCTGTTTGGGAG	CGCTTGGAGTGGAAA	CAAAAGGTGGGCAT	CAAAAGATGAGAAC	CCAAA	3420
Db	3361	CAGCCTGTTTGGGAG	CGCTTGGAGTGGAAA	CAAAAGGTGGGCAT	CAAAAGATGAGAAC	CCAAA	3420
QY	3421	GGCCCTACAACTCC	AGCCAGCCAGCCAG	GAGGGGTGTTC	CAATCAATCAGCAT	TTGGGA	3480
Db	3421	GGCCCTACAACTCC	AGCCAGCCAGCCAG	GAGGGGTGTTC	CAATCAATCAGCAT	TTGGGA	3480
QY	3481	ATGAGCTGGGCGCT	TGGGTGAGGTGGGG	GTCTGGCTTA	GTGGGAGGGCCCTG	GGGT	3540
Db	3481	ATGAGCTGGGCGCT	TGGGTGAGGTGGGG	GTCTGGCTTA	GTGGGAGGGCCCTG	GGGT	3540
QY	3541	GGGGGAGGGCGCT	TGGGTGAGGTGGGG	GTCTGGCTTA	GTGGGAGGGCCCTG	GGGT	3600
Db	3541	GGGGGAGGGCGCT	TGGGTGAGGTGGGG	GTCTGGCTTA	GTGGGAGGGCCCTG	GGGT	3600
QY	3601	AGGTCGTGATTGGG	GATGGGACAAAGAA	TAGCAAGAGATG	AGAAACACAGAA	AACTTT	3660
Db	3601	AGGTCGTGATTGGG	GATGGGACAAAGAA	TAGCAAGAGATG	AGAAACACAGAA	AACTTT	3660
QY	3661	TTTTCTTAAAGACT	GGTTAAATCAATTC	TGTATACAGCCTTA	CAATTAATAGATG	CAG	3720
Db	3661	TTTTCTTAAAGACT	GGTTAAATCAATTC	TGTATACAGCCTTA	CAATTAATAGATG	CAG	3720
QY	3721	CTAAAAATTAATTT	GTATGTCTTATAT	ACTATATATATAT	ATCTTCAGGTA	AAAAAGGC	3780
Db	3721	CTAAAAATTAATTT	GTATGTCTTATAT	ACTATATATATAT	ATCTTCAGGTA	AAAAAGGC	3780
QY	3781	AAGCCACAGAAAT	GTGTATAGGCGCAT	TCCATTTGTGTTC	CAGAAAGAGATGA	TATA	3840
Db	3781	AAGCCACAGAAAT	GTGTATAGGCGCAT	TCCATTTGTGTTC	CAGAAAGAGATGA	TATA	3840
QY	3841	AACACATTAATTT	GCTATATGCTATG	CCATTCAGAAATTA	ATGGGTAACAGTAT	ACTTTTGG	3900
Db	3841	AACACATTAATTT	GCTATATGCTATG	CCATTCAGAAATTA	ATGGGTAACAGTAT	ACTTTTGG	3900
QY	3901	GAGGGAAACGAT	AGTTGAGGACAG	GAGAGGGAAGG	GTCTTAACACTTAC	CCCTTTTG	3960
Db	3901	GAGGGAAACGAT	AGTTGAGGACAG	GAGAGGGAAGG	GTCTTAACACTTAC	CCCTTTTG	3960
QY	3961	TACATTTGAATTT	TGAACCATGTGAC	GTATTACCTTA	CAAAATTAACAATTA	ATGGG	4020
Db	3961	TACATTTGAATTT	TGAACCATGTGAC	GTATTACCTTA	CAAAATTAACAATTA	ATGGG	4020
QY	4021	CCCAAAAAAAAAAAAA	AAAAA	4040			
Db	4021	CCCAAAAAAAAAAAAA	AAAAA	4040			
RESULT 4							
LOCUS	AF067406		2501 bp	mRNA	Linear	PRI 14-NOV-2001	
DEFINITION	Homo sapiens vascular adhesion protein-1 (VAP1)			mRNA, complete cds			
ACCESSION	AF067406						
VERSION	AF067406.1						
KEYWORDS							
SOURCE							
ORGANISM	Homo sapiens.						
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.						
REFERENCE							
TITLE	1 (bases 1 to 2501)						
AUTHORS	Smith,D.J., Salmi,M., Bono,P., Hellman,J., Leu,T. and Jalkanen,S.						
	Cloning of vascular adhesion protein 1 reveals a novel						
	multifunctional adhesion molecule						
	J. Exp. Med. 188 (1), 17-27 (1998)						
JOURNAL	98317014						
MEDLINE	9653080						
PUBMED	2 (bases 1 to 2501)						
REFERENCE	Smith,D.J., Salmi,M., Bono,P., Hellman,J., Leu,T. and Jalkanen,S.						
AUTHORS	Direct Submission						

JOURNAL Submitted (20-MAY-1998) Biotite Therapies Ltd, Tylistokatu 6, Turku
FIN-20520, Finland

FEATURES

source Location/Qualifiers

1..2501

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17q21"
/tissue_type="lung"
1..2501

/gene="VAP1"
81..2372

/EC_number="1.4.3.6"
/note="semicarbazide sensitive amine oxidase;
copper-containing amine oxidase homology
/codon_start=1
/product="vascular adhesion protein-1"
/protein_id="M25170.1"
/db_xref="GI:3283370"

/translation="MNKDTLITLAVITFALVYLYVGRGDGGEPSQLPCHPSV
SPSQPWTTHPEQSLEPFDLSREELTAVMRFLTQRLGGLVDAQARSDNCVSVELQ
LIPFAALALDLRGSPPAPKALAVFFGSGPOPNSELVAGPLPHPSYMTATVERH
GGPLPHRRPLVLEQSEYLDIDQMLENRELQASGLAHNCEYKHGRNLVMTAPRL
QSGGRATVFGLYYINISGAGFELHHVGLLELVNKHALDPAKMTIDKYFYQGRYASLQ
LEAFEGALVNVVLIPLNGTGSWSLKSVPYGPAPLPLOYPGQRFSGVRSASLF
WTFSEFGAGSGRPIFDVRFGESELYEISLEELAIYGGNSPAAMTRVYDGGFPG
KMTPLRGVDCPYLATVYDMHFLLESQAKRTIDACVCEENQGLRLRHSHLYSH
YFGSLAETVLYVRMSSTLINTDYMDTVFHRSCIEIRLATGTSALFEGATGTG
NOVSEHTLVFHTHSAHFKVDLDVAGLENNVMAEDVFPMAVPMSEHOLDRLQVTR
KLEMEQDAFLVGSATPRYLALYLSNHSNMWGRGRYRIDLSPAGRLPONSWMAG
FSIMRYOLIAVORKEEPSSESVFNQNDMPAPVDFSEDTINNETIACKDLVAMTAF
LHMRAEDPNTVGVNGVGFELRPYNFEDPESFYSDSIFYRQDDADACEVNPFLA
CLPQAAACAPLDLPFSGHGFSHN"

BASE COUNT 464 a 782 c 719 g 536 t

ORIGIN

Query Match 61.9%; Score 2501; DB 9; Length 2501;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 81 GCCACAGAGCCCCCGTGTGGTGGCGTGAAGATTCATTGCTCCTTGGTGAATAG 140

Db 1 GCCAACAGAGCCCCCGTGTGGTGGCGTGAAGATTCATTGCTCCTTGGTGAATAG 60

0Y 141 CTGTCCCTTTCGTGGAAAAAGAACCAAGAACAAATCTGCTCCTCAATTCGGCG 200

Db 61 CTGTCCCTTTCGTGGAAAAAGAACCAAGAACAAATCTGCTCCTCAATTCGGCG 120

0Y 201 TCATACCATCTTTCCTTGGTTTGTGTCTTCCTGCTGGTGGCAAGGGGTGAGATGGGGTG 260

Db 121 TCATACCATCTTTCCTTGGTTTGTGTCTTCCTGCTGGTGGCAAGGGGTGAGATGGGGTG 180

0Y 261 AACCAGCAGGTTCCCATTTGCCATTTGCTGATCTCCCAAGGCCAGCTTGGACACAC 320

Db 181 AACCAGCAGGTTCCCATTTGCCATTTGCTGATCTCCCAAGGCCAGCTTGGACACAC 240

0Y 321 CTGGCCAGAGCCAGCTGTTTGCAGACTGAGCCGAGAGAGACTGAAGCTGTGATGGCCT 380

Db 241 CTGGCCAGAGCCAGCTGTTTGCAGACTGAGCCGAGAGAGACTGAAGCTGTGATGGCCT 300

0Y 381 TTCTACACCCAGCGGTGGGGGCCAGGGGCTGTGTGATGACGCCAGGCCCGGCCCTTGAC 440

Db 301 TTCTACACCCAGCGGTGGGGGCCAGGGGCTGTGTGATGACGCCAGGCCCGGCCCTTGAC 360

0Y 441 ACTGTGTCCTTCAGTGTGAGTGTGACGTGCGCTCCCAAGGCTGACAGCCCTGTGCTACTTGG 500

Db 361 ACTGTGTCCTTCAGTGTGAGTGTGACGTGCGCTCCCAAGGCTGACAGCCCTGTGCTACTTGG 420

0Y 501 ACAGGGGAGCCCCCACCCTGCGGGGAGGACATGCGCATGCTCTTTTGGACAGCAAC 560

Db 421 ACAGGGGAGCCCCCACCCTGCGGGGAGGACATGCGCATGCTCTTTTGGACAGCAAC 480

Oy	561	CCGACGCCAACGTAGTGAATGGTGTGGTGCGCATGGCTCAACCCCTCTACTATGGGG	620
Db	481	CCCCAACCACGTAGTAGGAGCTGTGTGGGGCAGCTGCCTCACCCCCTCTACATGGGG	540
Oy	621	ACGTGACTGTGGAGCGTCATGGAGGGCCCTGCGCTATCCAGAGCGCCCGCTGTTC	680
Db	541	ACGTGACTGTGGAGCGTCATGGAGGGCCCTGCGCTATCCAGAGCGCCCGCTGTTC	600
Oy	681	AAGGTACTGTGACATAGACAGATGATCTTTCAAACAGAGAGTGGCCCAAGGCTTGGGC	740
Db	601	AAGGTACTGTGACATAGACAGATGATCTTTCAAACAGAGAGTGGCCCAAGGCTTGGGC	660
Oy	741	TTCCTCACCAGCTGTGCTTCTTACAAGCACCAGGAGGAACTCTGTGACATGACACGG	800
Db	661	TTCCTCACCAGCTGTGCTTCTTACAAGCACCAGGAGGAACTCTGTGACATGACACGG	720
Oy	801	CTCCCGGCGGTCTGACATGAGGGAGCGGGCAGCTGGTTGGCGCTCTACTACATGCT	860
Db	721	CTCCCGGCGGTCTGACATGAGGGAGCGGGCAGCTGGTTGGCGCTCTACTACATGCT	780
Oy	861	CAGGCGCTGGGTTCTTCCTGCACACGCTGGGCTTGGAGCTGTAAGTAAACACAGGCC	920
Db	781	CAGGCGCTGGGTTCTTCCTGCACACGCTGGGCTTGGAGCTGTAAGTAAACACAGGCC	840
Oy	921	TTGACCTGCCCCGCTGGACATCTCCAGAAGTGTCTTATCAAGCCGCTACTACGACAGC	980
Db	841	TTGACCTGCCCCGCTGGACATCTCCAGAAGTGTCTTATCAAGCCGCTACTACGACAGC	900
Oy	981	TGGCCACCTGGAAGGCCAGTTTGAAGCGGGCGGTGAATGTGGTCTATCCAGACA	1040
Db	901	TGGCCACCTGGAAGGCCAGTTTGAAGCGGGCGGTGAATGTGGTCTATCCAGACA	960
Oy	1041	ATGGCAGAGTGGGTCTCTGTGCTCCCTGAAATCCCTGTGCCCCCGGGTCCAGCTCCCTC	1100
Db	961	ATGGCAGAGTGGGTCTCTGTGCTCCCTGAAATCCCTGTGCCCCCGGGTCCAGCTCCCTC	1020
Oy	1101	TACAGTCTTATCCCAAGAGGCCCGCGCTTCACTGTCCAGGGAAGTCCAGTGGCTCTCAC	1160
Db	1021	TACAGTCTTATCCCAAGAGGCCCGCGCTTCACTGTCCAGGGAAGTCCAGTGGCTCTCAC	1080
Oy	1161	TGTGACTTTCCTCTTGGGCTCGGAGCATTCAGTGGCCCAAGGATTTTGAGCTTGCT	1220
Db	1081	TGTGACTTTCCTCTTGGGCTCGGAGCATTCAGTGGCCCAAGGATTTTGAGCTTGCT	1140
Oy	1221	TCCAAGGAGAAAGATATTATGAGATAGCTTCACAAGAGGCTTGGCCATCATGGTG	1280
Db	1141	TCCAAGGAGAAAGATATTATGAGATAGCTTCACAAGAGGCTTGGCCATCATGGTG	1200
Oy	1281	GAAATTTCCCAAGACCAATGACAGACCGGCTATGTGATGAGGCTTTGGCATGGGCAAT	1340
Db	1201	GAAATTTCCCAAGACCAATGACAGACCGGCTATGTGATGAGGCTTTGGCATGGGCAAT	1260
Oy	1341	ACACACGCGCCCTGACCGTGGGGGAGACTGCCCTACTTGGACACCTAGCTGGAGCTGGC	1400
Db	1261	ACACACGCGCCCTGACCGTGGGGGAGACTGCCCTACTTGGACACCTAGCTGGAGCTGGC	1320
Oy	1401	ACTTCTTTTGGAGTCCAGAGCGCCCAAGACCAATACGTGATGCTTTGTGTGTTTAAAC	1460
Db	1321	ACTTCTTTTGGAGTCCAGAGCGCCCAAGACCAATACGTGATGCTTTGTGTGTTTAAAC	1380
Oy	1461	AGACCCAGAGGCTCCCTCGTGGGGGACACACACTCATGATCTTACTCGCATCTTTGGGG	1520
Db	1381	AGACCCAGAGGCTCCCTCGTGGGGGACACACACTCATGATCTTACTCGCATCTTTGGGG	1440
Oy	1521	GTCTTGGGAAACGGTGGTGGTGTGATGATGTGATGCTGCTCAACTGATGATG	1580
Db	1441	GTCTTGGGAAACGGTGGTGGTGTGATGATGTGATGCTGCTCAACTGATGATG	1500
Oy	1581	TGTGGGATACGGTCTTCCACCCCACTGGGGCATAGAAATACGAATTTGATGCCAGGGCT	1640
Db	1501	TGTGGGATACGGTCTTCCACCCCACTGGGGCATAGAAATACGAATTTGATGCCAGGGCT	1560
Oy	1641	ACATCAGCTGGCATCTCTCTTGTGTGCTACTGGGAGTACGGGAACCACTGTCCAGNC	1700
Db	1561	ACATCAGCTGGCATCTCTCTTGTGTGCTACTGGGAGTACGGGAACCACTGTCCAGNC	1620
Oy	1701	ACACCTTGGGACAGCTGCACACACCACAGCGCCCACTTCAAGTGTGATGTAGAC	1760
Db	1621	ACACCTTGGGACAGCTGCACACACCACAGCGCCCACTTCAAGTGTGATGTAGAC	1680
Oy	1761	GACTGGAAATGTGGTCTGGGCGGAGATATGTGCTTTTGTCCCATAGCTGTGCCCTGA	1820
Db	1681	GACTGGAAATGTGGTCTGGGCGGAGATATGTGCTTTTGTCCCATAGCTGTGCCCTGA	1740
Oy	1821	GCCCTGAGCAACAGCTGCAGAGGCTGAGGTGACCCGGAAGTGTGTGATGTAGAGAC	1880
Db	1741	GCCCTGAGCAACAGCTGCAGAGGCTGAGGTGACCCGGAAGTGTGTGATGTAGAGAC	1800
Oy	1881	AGGCGGCTTCTCTGTGGGAAGCGCCACCCCTGCTACCTGTACTGGCCAGCAACACA	1940
Db	1801	AGGCGGCTTCTCTGTGGGAAGCGCCACCCCTGCTACCTGTACTGGCCAGCAACACA	1860
Oy	1941	GCAACAAATGGGGTCAACCCCGGGGCTACCGCATCCAGATGCTCAGCTTGTGGAGAC	2000
Db	1861	GCAACAAATGGGGTCAACCCCGGGGCTACCGCATCCAGATGCTCAGCTTGTGGAGAC	1920
Oy	2001	CGCTGCCCAAAACAGCTCCATGGCGAGAGGCTTCAAGCTGAGGAGGTACACAGCTG	2060
Db	1921	CGCTGCCCAAAACAGCTCCATGGCGAGAGGCTTCAAGCTGAGGAGGTACACAGCTG	1980
Oy	2061	TGACCAGCGGAAGAGAGAGGAGCCAGTAGACAGACCGTTTTCAATCAGATGACCTT	2120
Db	1981	TGACCAGCGGAAGAGAGAGGAGCCAGTAGACAGACCGTTTTCAATCAGATGACCTT	2040
Oy	2121	GGGCCCCCACTGTGGATTTCACTAGTACTTATCAACAAATGACACATTTGCTGAAAAGATT	2180
Db	2041	GGGCCCCCACTGTGGATTTCACTAGTACTTATCAACAAATGACACATTTGCTGAAAAGATT	2100
Oy	2181	TGTGGGCTGGGTGACAGCTGGTTTTCGATATCCACATGCGAGAGACATTCCTAAC	2240
Db	2101	TGTGGGCTGGGTGACAGCTGGTTTTCGATATCCACATGCGAGAGACATTCCTAAC	2160
Oy	2241	CAGTACTGTGGGGAACGGCTGGGCTTCTCTCCAGCCCTATTAACCTTTTGAGAGAG	2300
Db	2161	CAGTACTGTGGGGAACGGGCTGGGCTTCTCTCCAGCCCTATTAACCTTTTGAGAGAG	2220
Oy	2301	ACCCTCTTCTACTCTCGCAGCTCATCTACTTCCAGAGGAGACAGATGCTGGGGCT	2360
Db	2221	ACCCTCTTCTACTCTCGCAGCTCATCTACTTCCAGAGGAGACAGATGCTGGGGCT	2280
Oy	2361	GCGAGGTCAACCCCTACTTGGCTGCCCGCAGGCTGTGGCTGTGCCCGGACCTCGTG	2420
Db	2281	GCGAGGTCAACCCCTACTTGGCTGCCCGCAGGCTGTGGCTGTGCCCGGACCTCGTG	2340
Oy	2421	CTTCTCCACAGGGGGCTTCTCTCAACAATAGGGGCTCTTGGATGGGGCATGTGGCAA	2480
Db	2341	CTTCTCCACAGGGGGCTTCTCTCAACAATAGGGGCTCTTGGATGGGGCATGTGGCAA	2400
Oy	2481	GGGCTCCAGGGCGCAGGCTGTGAGGATGGGAGACGCTGGGCACTGGGCGGACCTGG	2540
Db	2401	GGGCTCCAGGGCGCAGGCTGTGAGGATGGGAGACGCTGGGCA	

ORGANISM clone:HEP11945.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Kawabata,A., Hiki,J., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEO human cDNA sequencing project
JOURNAL
REFERENCE 2 (bases 1 to 2538)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL
COMMENT Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
Japan (E-mail:cdna@leims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
NEO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5' - 3' -end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
Source 1. .2538
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HEP11945"
 /cell_line="HepG2"
 /cell_type="hepatoma"
 /clone_id="HEP"
 /note="Cloning vector pME18SFL3"
 1. .2538
 /note="Highly similar to HSU39447 Human placenta copper
 monamine oxidase mRNA"
BASE COUNT 554 a 687 c 696 g 601 t
ORIGIN
Query Match 56.4%; Score 2277; DB 9; Length 2538;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2280; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2176 GCATTGGTGGCTGGGTGACAGCTGCTGTTTCTGCATATCCCATGAGAGACATTCC 2235
 |||||
DB 666 GGAATTGGGGCTGGGTGACAGCTGCTGTTTCTGCATATCCCATGAGAGACATTCC 725
QY 2236 TAACAGAGTATGGGGGAAGGGGTGGGGCTTCTCCGACCCCTAATCTTTGA 2295
 |||||
DB 726 TAACAGAGTATGGGGGAAGGGGTGGGGCTTCTCCGACCCCTAATCTTTGA 785
QY 2296 CGAAGACCCCTCTCTACTCTGCGACTCATCTACTTCCAGGGGACAGAGATGTGG 2355
 |||||
DB 786 CGAAGACCCCTCTCTACTCTGCGACTCATCTACTTCCAGGGGACAGAGATGTGG 845
QY 2356 GGCCTGCGAGTCAACCCCTAGCTTGCCTGCCAGCGCTGCTGCTGCCGACCT 2415
 |||||
DB 846 GGCCTGCGAGTCAACCCCTAGCTTGCCTGCCAGCGCTGCTGCTGCCGACCT 905
QY 2416 CCTGCTTCTCCACGCGGGGCTTCTCACAACATAGCGGTCGAGGATGGGGATGTG 2475
 |||||
DB 906 CCTGCTTCTCCACGCGGGGCTTCTCACAACATAGCGGTCGAGGATGGGGATGTG 965
QY 2476 GCCAAGGCTCCAGGGCCAGAGGTGTGAGGATGGGAGACAGCTGGGACATGGCCGAG 2535
 |||||
DB 966 GCCAAGGCTCCAGGGCCAGAGGTGTGAGGATGGGAGACAGCTGGGACATGGCCGAG 1025
QY 2536 CCTGCTTCTCTCTCTGTCGACAGACTCTTCTTCTCAGTACCTCCCTCGCATCCG 2595
 |||||
DB 1026 CCTGCTTCTCTCTCTGTCGACAGACTCTTCTTCTCAGTACCTCCCTCGCATCCG 1085
QY 2596 CCTGTAGACAGAGACCTCTGACCTGATGATGCTGACACAGGGGACATGAACCTGT 2655
 |||||
DB 1086 CCTGTAGACAGAGACCTCTGACCTGATGATGCTGACACAGGGGACATGAACCTGT 1145
QY 2656 TGATCCAGCTGACTGAGTTCTCATCCACAGAGCCAGAGCATGGCCAGCTGGAGCG 2715
 |||||
DB 1146 TGATCCAGCTGACTGAGTTCTCATCCACAGAGCCAGAGCATGGCCAGCTGGAGCG 1205
QY 2716 TGGCGAGAGGCTTCCCTAGATGATGCTTCTGCTGTGCTTCCGAAATTTT 2775
 |||||
DB 1206 TGGCGAGAGGCTTCCCTAGATGATGCTTCTGCTGTGCTTCCGAAATTTT 1265
QY 2776 AGGCCACTCCAAAGACCTTAAAGGGGCTATTCCCTGAGAGCCCAAGAGAGGTTGC 2835
 |||||
DB 1266 AGGCCACTCCAAAGACCTTAAAGGGGCTATTCCCTGAGAGCCCAAGAGAGGTTGC 1325
QY 2836 CAGTCCCTCAAGTCAATAGCTGAGCTGGAAGAGATGCTTGTGCATTCCTCTATC 2895
 |||||
DB 1326 CAGTCCCTCAAGTCAATAGCTGAGCTGGAAGAGATGCTTGTGCATTCCTCTATC 1385
QY 2896 CAGTCCCT 2955
 |||||
DB 1386 CAGTCCCT 1445
QY 2956 TGCCT 3015
 |||||
DB 1446 TGCCT 1505
QY 3016 CCCCTGTACTCTCCAGCCCTCACTTTTCAAGTCAAGTCCGCTCTCTCTCTCTCTCTCT 3075
 |||||
DB 1506 CCCCTGTACTCTCCAGCCCTCACTTTTCAAGTCAAGTCCGCTCTCTCTCTCTCTCT 1565
QY 3076 GGAAGTCTCAAGGTACAGGGACCCCTAATCAAGTGGCCATTCCTGTGTGCTTCCCT 3135
 |||||
DB 1566 GGAAGTCTCAAGGTACAGGGACCCCTAATCAAGTGGCCATTCCTGTGTGCTTCCCT 1625
QY 3136 TGTGTCTGTGCTTATTGGAGTAGAGTGTCTCTACCCCTGTCTGGGGTGGGTGTG 3195
 |||||
DB 1636 TGTGTCTGTGCTTATTGGAGTAGAGTGTCTCTACCCCTGTCTGGGGTGGGTGTG 1685
QY 3196 TTTCAAGACAGCTGCTTGTGTGATTTGTGTCTGCTCTCTCTCTCTCTCTCTCTCT 3255
 |||||
DB 1686 TTTCAAGACAGCTGCTTGTGTGATTTGTGTCTGCTCTCTCTCTCTCTCTCTCTCT 1745
QY 3256 GATGTCATCTGTGACAGCAGCTCAAGTTAGCATTTTCAAGTATTTGGGGGTGCAATG 3315

Db	400	GTGATGAGCTTCCTGACCCAGAACAGTGGGGCCAGACCTGGTGGATGGAGCCAGGGCCGA	459
Oy	431	CCCTGGGCAACATGTGCTTCTCAGTGGAGTTGCAGCTGCTCCCAAGGCTGGAGCCCTG	490
Db	480	CCCTCTGACACATGCATCTTCTTGCGTAGAGCTGAGCTGCCCCCAAGGCTGCAAGCCCTG	519
Oy	491	GCTCACTTGGACAGAGGGGAGCCCCCACACCTGCCGGGAGGAGGACTGGCCATGCTTCTTT	550
Db	520	GCCCAACCTGGACAGAGGAGAGCCCCCACCTGCCGGGAGGAGGACTGGCCATGCTTCTTT	579
Oy	551	GGCAGGCAACCCCGAGCCCAACGTGATGAGACTGGTGGGGGCCACTGGCCATGCTTCTTT	610
Db	580	GGTGGACAAACCCCGAGCCCAATGATACGAGCTGGTGGTGGGGCCGCTGCCAGCCCTCC	639
Oy	611	TACATCGGGGAGCTGAGACTGTGAGAGCGTCATAGAGAGCCCCCTGCCCTATCCAGAACGCC	670
Db	640	TACATCGCGGAGTGTGACCCGTGAGAGCGTCATAGGGGGCCCCCTGCCCTATTCAGCAAGCCC	699
Oy	671	GTGCTGTCCAGAGAGTACCTGGACATAGACAGATGATCTTCAACAGAGAGAGTGGCCCA	730
Db	700	GTGCTTCTCCGAGAGTACCTGGACATAGACAGATGATCTTCAACAGAGAGTGGCCCA	759
Oy	731	GCTTCTGGGCTTCTCCACACATGTGCTTCAACAGACCGGGGAGGAACTGGTGACA	790
Db	760	GCTCTGGTGTCCGACACATGTGCTCTCAACAAAGAGAGAGGAACTGGTGACC	819
Oy	791	ATGACACAGGCTCCCGCTGGCTCAATCAAGGGAGCCGGGCCACAGTGGTGGCCCTAC	850
Db	820	ATGACACAGAGCCCCCGCGGTTTGCAATCAAGGTACCGGGCCACCTGCTGGACCTCTAC	879
Oy	851	TACAACATCTCGGGGCGGTGGGTTCTTCTCAGCACACGTGGGCTTGGAGCTGTAGTAA	910
Db	880	TACAACATCTCAGGGGCGGTGGTACTATCTGACACCTGTGGGGTTAGAGCTGTGTAGAT	939
Oy	911	CACAAGGCGCTTGACCTGCCCGCTGGACATATCCAGAAGTGTTCATCAAGGCCCGCTAC	970
Db	940	CACAAGGCTCTGGACCTCGCCCAATGAGACCATCCAGAAGGTGTCTTCAAGGCCCGCTAC	999
Oy	971	TACGACAGCCCTGGGCCAGAGCTGGAGAGCCCAATTTGAGGCCGGCGCTGGTGAATGTGGTCTG	103
Db	1000	TATGAAAGTCTGGCCACAGCTGAGAGAGCATTTGAGCTGGCCGGGTGAATGTGGTGTG	105
Oy	1031	ATCCGAGACAAATGGCACAGAGTGGGCTCTGGTCCCTGAAGTCCCTGTGCCCCCGGGTCCA	109
Db	1060	ATCCCAAAACAAATGGGACAGGAGGGTCTCTGGTCCCTGAAGTCCCAAGTGGCCCCGGGTCCA	111
Oy	1091	GCTGCCCTCTACAGTTCATCCCCAAGGCCCCCGCTTCAAGTGTCCAGGAACTCGAGTG	115
Db	1120	ACTCCACTCTGCAAGTTCATCCATCCCAAGGACCCGCTTCAAGTGTCCAGGCACTCGAGTG	117
Oy	1151	GCGTCTACAGTGGAGACTTCTCCTTTGGGCGCTGGAGCATTCAGTAGTGGCCCCAAGATCTTT	121
Db	1180	ACCTCTCATATGTGACTTCTCTCTTTGGCGCTGGAGCTTTCAGTGTCTCTAGATCTTTT	123
Oy	1211	GACGTTGCTTCCAAAGGAGAAAGCTAGTTATGATAGACTCTCCAAAGAGGCGTTGGCC	127
Db	1240	GACATTTGATTTCCAAAGGAGAAAGACTGGCTTATGAGATCAAGCTCGAAGAGGCGGTGGCT	129
Oy	1271	ATCTATGTGGAAATTTCCCAAGCAGCAATGACGCCGTATGTGATGGATGGAGGCTTTGGC	133
Db	1300	ATTATATGTGGAAATTTCCCAAGCAGCAATGCTACTCTGCTATATGATGATGGCTCTTTGGC	135
Oy	1331	ATGGGCAAGTACACACAGGCGCCGTGACCCGAGGGGTGAGACGCCCTTACTTGGCCACTAC	139
Db	1360	ATGGGCAAGTTCGCGACAGCCCGCTGACCCGAGGGGTGAGACGCCCTTATCTGGCCACTAT	141
Oy	1391	GTGAGACTGGCACTTCTTTTGGAGATCCCAAGGCCCCCAAGACAAATAGTGAATGGCTTTGT	145
Db	1420	GTGAGACTGGCACTTCTTCTTGGAGATCTCAAGCCCCCAAGACCTACATGATGCTTTTGT	147
Oy	1451	GTTGTTGAACAGAACAGGGGCTTCCCCGTGGGAGACCACTCAGATCTCTACTGGCAC	151

[illegible]

2591 ATCCGCTGTAGCCAGGAGCCCTCTGACCTGTATGCTGACAGAGGGACCTGAAAC 2650
 2620 GCTTACCGGCTCACTGAGGATCATCGTATGCTGTAATAGAGGGCTTCTCAGC 2679
 2651 CTTGTGATGACGAGCTGTACTGAGTGTCTATCCACAGAGG---CCAGGATGGCCAGC 2706
 2680 TCTGTGACACCCCACTTGTGCTGCTTCTGTCGAGAGAGAGAGAGGATGACCAAGAGG 2739
 2707 CTGAGCCGCTGGCCGAGGCTTC-----CCTAGATG 2737
 2740 CTGAGACTCATGTTTACAGCTTCTACAGACTCTCTGCTTGTGTTTTCATTTCCTT 2799
 2738 GTTCCCTTGTGCTGCTGCTGCTTCCCAATTTTATAGCCACTCCAGAGACTCTA 2797
 2800 TCTTTGTCTTGTCTTGTCTTGTCTTCCCAATGTTTATAGGCGCATGCGAGTCTCTG 2859
 2798 AAGGGGCTATTCCCTGAGAGACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2857
 2860 AGACTCCCACTCTCACTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2919
 2858 AGCTGAAAG 2917
 2920 AGAGGCGGAG 2979
 2918 CTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2977
 2980 TCTTTGCGGCTATTTCT 3039
 2978 TTTCTCCGGAATCTGAGGGAGATATCCCTATGTCGCCAGCCCTGATGCTCCCGAG 3037
 3040 CT 3097
 3038 CAGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3097
 3098 TGTTCCTCAGCTCTCACTTCTTCACTGCTGCTCTCTCTCTCTCTCTCTCTCTCT 3156
 3098 CCTTATCAG 3157
 3157 GTGTGAATATGAG 3216
 3158 TAGAGAGTGTCT 3217
 3217 TGGAGAGCGGCGCT 3275
 3218 CATTGTGTCT 3277
 3276 CATTGTGTATGAGGCTG--TCTTGCATCTCTCTCTCTCTCTCTCTCTCTCTCT 3333
 3278 GCTCAAGTTTGAATTCATGATTTGGGGTGTGATGATGATGATGATGATGATGAT 3337
 3334 GCTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3393
 3338 TACAGAGGCTCTGATTTCTCTCAACAGAGCTTTTGGAGGCTGAGTGGAA--CA 3392
 3394 TGCAGAGT 3453
 3393 GGGTGGGATCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3452
 3454 GGGGAAAG 3511
 3453 GGGTGGGATCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3512
 3512 GGGGAG 3571
 3513 GCGTAG 3518
 3572 CCACAG 3577

RESULT 7
 AC046171/c
 LOCUS AC046171 189149 bp DNA 1linear HTG 03-MAR-2001

DEFINITION Homo sapiens chromosome 17 clone RP11-948G15 map 17, WORKING DRAFT
 AC046171
 VERSION AC046171.3 GI:13184209
 KEYWORDS HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 189149)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 17, clone RP11-948G15
 Unpublished
 2 (bases 1 to 189149)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bida, F.,
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
 Campopiano, L., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Deatrelano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferrerella, P., Fitzhugh, N., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McNeesters, R.,
 Melidim, J., Menus, L., Mihova, T., Miranda, C., Mienga, Y., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Ollivar, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange, R., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testaye, S., Theodore, J., Tirrell, A., Travers, M., Tsigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (13-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 2, 2001 this sequence version replaced g1:7712190.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L7946
 Center clone name: 948.G.15
 ----- Summary Statistics
 Sequencing vector: M13; M7815; 408 of reads
 Sequencing vector: Plasmid; n/a; 608 of reads
 Chemistry: Dye-terminator; Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 185448 bases at least Q40
 Consensus quality: 186995 bases at least Q30
 Consensus quality: 187536 bases at least Q20
 Insert size: 190000; agarose-fp
 Insert size: 187849; sum-of-contigs
 Quality coverage: 10.0 in Q20 bases; agarose-fp
 Quality coverage: 10.1 in Q20.
 NOTE: This is a 'working draft' sequence. It currently
 consists of 14 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.
 1 1116: contig of 1116 bp in length
 1117 11216: gap of 100 bp
 11217 13252: contig of 2036 bp in length

Source

misc_feature

misc_feature

ORIGIN

Match

Dd	1612350	AGGATTTGGTGGCCCTGGGTGACAGCTGGTATTTCGATATTCACATATGACAGAGCAATTC	162191
Qy	2235	CTAACACAGTAGACTGTGGGGGAAGAGGCGGGGCTTCCTCTCTCCAGCCCTATTAACCTCTTGG	2294
Dd	161230	CTAACACAGTAGACTGTGGGGGAAGAGGCGGGGCTTCCTCTCTCCAGCCCTATTAACCTCTTGG	162131
Qy	2295	ACGAAGACCCCTCTTCTACTCTGCGCATCTCATCTTCCAGGGAGACAGAGATCTG	2354
Dd	162130	ACGAAGACCCCTCTTCTACTCTGCGCATCTCATCTTCCAGGGAGACAGAGATCTG	162120
Qy	2355	GAGGCTGGAGAGTCAACCCCTACCTTGGCTGGCCCAAGGCTGCTGCTGGCCCCGACC	2414
Dd	162070	GAGGCTGGAGAGTCAACCCCTACCTTGGCTGGCCCAAGGCTGCTGCTGGCCCCGACC	162011
Qy	2415	TCCCGGCTTCCACAGGGGGGCTTCTCTCAACTAGGGGCTCTGGATGGGGCAATG	2474
Dd	162010	TCCCGGCTTCTCCACAGGGGGCTTCTCTCAACTAGGGGCTCTGGATGGGGCAATG	161951
Qy	2475	GGCCAAGGGCTCCAAGGGCCAGAGGTGTAGAGATGGGAGACAGTGGGACATGGGCGCA	2534
Dd	161950	GGCCAAGGGCTCCAAGGGCCAGAGGTGTAGAGATGGGAGACAGTGGGACATGGGCGCA	161899
Qy	2535	GCCCTGTCCTCTTCTTCTGTGCCAGAGACTCTTCTTCTTCCACTACCTCCCTGCAATC	2594
Dd	161890	GCCCTGTCCTCTTCTTCTGTGCCAGAGACTCTTCTTCTTCCACTACCTCCCTGCAATC	161833
Qy	2595	GCCCTGAGCCAGAGAGGCTCCTGACCCGTGATGCTCGACACAGGGGACACTGAACCTTG	2654
Dd	161830	GCCCTGAGCCAGAGAGGCTCCTGACCCGTGATGCTCGACACAGGGGACACTGAACCTTG	161777
Qy	2655	TTGATGCCAGCTGACTGATGTTCTCATCTCACAGAGGCGAGGCAATGGCCACGCGAGACC	2714
Dd	161770	TTGATGCCAGCTGACTGATGTTCTCATCTCACAGAGGCGAGGCAATGGCCACGCGAGACC	161711
Qy	2715	GTGGCCGAGGGCTTCCTAGATGGTTCCTTGTGTCTGTCTGGCTTCCCGAATCTTTT	2774
Dd	161710	GTGGCCGAGGGCTTCCTAGATGGTTCCTTGTGTCTGTCTGGCTTCCCGAATCTTTT	161653
Qy	2775	TAGGCGACCTCCAAAGGACTCTAAAGGGGGCTATTCCTGTGAGACCCCAAGATAGGGTTG	2834
Dd	161650	TAGGCGACCTCCAAAGGACTCTAAAGGGGGCTATTCCTGTGAGACCCCAAGATAGGGTTG	161591
Qy	2835	CCAGTCTGCAAGTCCATAGCTGAGCTGGAAGAGATGCTTGTGTCACATTTCCCTCAT	2894
Dd	161590	CCAGTCTGCAAGTCCATAGCTGAGCTGGAAGAGATGCTTGTGTCACATTTCCCTCAT	161533
Qy	2895	CCAGGCTCTTTCCTCTCTGCTCTCTCTCTCACTACTTCTCTCTCTCTCTGTTTC	2954
Dd	161530	CCAGGCTCTTTCCTCTCTGCTCTCTCTCTCTCACTACTTCTCTCTCTCTCTGTTTC	161471
Qy	2955	CTGCTTCTCTTCTTATCTGTGAATTTCTCCCGAATCCTGAGGGGATATCCCTATGTCCCA	3014
Dd	161470	CTGCTTCTCTTCTTATCTGTGAATTTCTCCCGAATCCTGAGGGGATATCCCTATGTCCCA	161411
Qy	3015	GCCCCGTGACTCCCCCAGGCCCTAGTTTTCAGTCAAGATTCCGCTACTCTCCAGCCCTA	3074
Dd	161410	GCCCCGTGACTCCCCCAGGCCCTAGTTTTCAGTCAAGATTCCGCTACTCTCCAGCCCTA	161351
Qy	3075	TGGAAGTCTCAAGGTACAGGGAGCCCTATATCAGATGAGGCCAATCCCTGTGTGCTTCCC	3134
Dd	161350	TGGAAGTCTCAAGGTACAGGGAGCCCTATATCAGATGAGGCCAATCCCTGTGTGCTTCCC	161291
Qy	3135	TTTGCTCTGTTGCTTATTTGGGAGTAGAGATTGCTCAACCCGTGTCTCTGGGGCGGGGTGT	3194
Dd	161290	TTTGCTCTGTTGCTTATTTGGGAGTAGAGATTGCTCAACCCGTGTGTCTCTGGGGCGGGGTGT	161231
Qy	3195	GTTCCTAGAGACAGCTGCTTCTGTGATTTGTGTGCTCTGCTCTCATGCTCTATATAGAGA	3254
Dd	161230	GTTCCTAGAGACAGCTGCTTCTGTGATTTGTGTGCTCTGCTCTCATGCTCTATATAGAGA	161171
Qy	3255	GGATGTCTATCTGTACAGACAGCAGCTCAAGTTACATTTCAAGTGAATTTGGGGTGCAT	3314
Dd	161170	GGATGTCTATCTGTACAGACAGCAGCTCAAGTTACATTTCAAGTGAATTTGGGGTGCAT	161111

DEFINITION Homo sapiens AOC3 gene for vascular adhesion protein-1, complete cds, exons 3 to 4.

ACCESSION AB050502

VERSION AB050502.1 GI:11602472

KEYWORDS

SEGMENT

SOURCE 3 of 3

ORGANISM Homo sapiens pre-pro-B cell cell_line:FLB14-14 DNA, clone_lib:keio BAC library clone:KB51F.

REFERENCE 1 Homo sapiens

AUTHORS Mammalia; Euteleostomi; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Primates; Carnivora; Homnidae; Homo.

TITLE Tamura, Y., Kudoh, Y., Mashima, Y., Oguchi, Y. and Shimizu, N. Tandem array of two human copper amine oxidase genes (AOC2 and AOC3) in 17q21

JOURNAL Unpublished

AUTHORS 2 (bases 1 to 3233)

TITLE Direct Submission

JOURNAL Submitted (25-OCT-2000) Nobuyoshi Shimizu, Keio University School of Medicine, Department of Molecular Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo 160, Japan (E-mail:shimizue@med.keio.ac.jp, Tel:03-3351-2370, Fax:03-3351-2370)

FEATURES

source

1..3233 location/qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="17"

/map="17q21"

/clone="KB51F"

/cell_line="FLB14-14"

/cell_type="pre-pro-B cell"

/clone_lib="Keio BAC library"

join(AB050500.1:639..2238,AB050501.1:581..866,556..685,1387..1662)

/gene="AOC3"

join(AB050500.1:639..2238,AB050501.1:581..866,556..685,1387..1662)

/EC_number="1.4.3.6"

/note="homolog of semicarbazide-sensitive amine oxidase placental monamine oxidase"

/codon_start=1

/product="Vascular adhesion protein-1"

/protein_id="BAB18866.1"

/db_xref="GI:11602473"

/translation="MNKTIIVLILAVITITFALVCVLLVGRGSGPSPHPCSVSPSAPWTHPGSOLFADLSREELTAVNRFORPGLVDVAQARPSDNVSEILOLPKRAALAHIDRSGPPARALAVFGRPPQPVSVSLVGRPLRPHRYMDVYERHGGPLRHRPVLFOEDLIDIDITRELPONSGLHNCCTKHNRNLVTMTARGLQSGRATWEGLYNISGAGFLVHAGLELVNKKALDPAWTTIOVFQGRYDSLAQLEQFAGLVNVLIPDNGTGSWSLSPVPPAPLPQFPOGRFVQGRVSLWTFSGAFSGPRIFDVRFOGSRLEYEISQELALVIGNSPAMTTRVYDGGGMSKTPPLTRGVDCPYLATVVDHNFLEISQAPRTIRAPCFVRDONGRLRRHNSDIYSHYEGGLAEVLYVRSSTLNDYWDYVFNHSGALEIFATGYTSSAFIGATKYGKQVSEHTLGIYHTSHAFKVDLDVAGLENNWVAEDWVEVPAVFPSPHOLDRLDYTRKLEMEQAAFLVGSATPRYLILASNSKNKGHPGRIYPMASFAGEPLPONSVARGFMSERYQLAVTORKEEPPSSSVFNONPAPVDVDFSDFINNETIAGDLVAMVWGLAIPIAEDIPNTVYVNGVGFELRPYNFDEDPSPYSADSIYFRDQDAGACEVNPPLCLPQAACAPDLPAPFSGHGSFN"

<1..555

/gene="AOC3"

556..685

/gene="AOC3"

/number=3

686..1386

/gene="AOC3"

1387..3233

/number=4

694 a 844 c 854 g 841 t

BASE COUNT

ORIGIN

Query Match 45.8%; Score 1849; DB 9; Length 3233;

Best Local Similarity 100.0%; Pred. No. 0; Matches 1849; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2175 AGGATTGGTGCCCTGGGTGACAGCTGGTTCATATCCACATGCAGAGACATTC 2234

Db 1385 AGGATTGGTGCCCTGGGTGACAGCTGGTTCATATCCACATGCAGAGACATTC 1444

QY 2235 CTACACAGTACTGTGGGGAACGGCTGGGCTTTCTCCCGAACCTTAATCTTTG 2294

Db 1445 CTACACAGTACTGTGGGGAACGGCTGGGCTTTCTCCCGAACCTTAATCTTTG 1504

QY 2295 ACGAAGACCCCTCTTACTGTGCGACTCATCTACTTCCGAGGGGACGATGTCG 2354

Db 1505 ACGAAGACCCCTCTTACTGTGCGACTCATCTACTTCCGAGGGGACGATGTCG 1564

QY 2355 GGGCTGCGAGGTCAACCCCTAGCTTGCTGCCCCAGGCTGCTGTGCCCGACC 2414

Db 1565 GGGCTGCGAGGTCAACCCCTAGCTTGCTGCCCCAGGCTGCTGTGCCCGACC 1624

QY 2415 TCCCTGCTTCTCCACAGGGGCTTCTCACACTAGGCGCTCTGGATGGGCATGT 2474

Db 1625 TCCCTGCTTCTCCACAGGGGCTTCTCACACTAGGCGCTCTGGATGGGCATGT 1684

QY 2475 GGGCAAGGCTCCAGGGCCAGGGTGTGAGGGATGGGGACACGFGGGCACTGGCGGCA 2534

Db 1685 GGGCAAGGCTCCAGGGCCAGGGTGTGAGGGATGGGGACACGFGGGCACTGGCGGCA 1744

QY 2535 GCGTGGTCCCTCTTCTCTGCTGCGACAGACTCTTCTTCTCACTACCTCCCTGCATCC 2594

Db 1745 GCGTGGTCCCTCTTCTCTGCTGCGACAGACTCTTCTTCTCACTACCTCCCTGCATCC 1804

QY 2595 GCGTGGTCCCTCTTCTCTGCTGCGACAGACTCTTCTTCTCACTACCTCCCTGCATCC 2654

Db 1805 GCGTGGTCCCTCTTCTCTGCTGCGACAGACTCTTCTTCTCACTACCTCCCTGCATCC 1864

QY 2655 TTGATGCCAGCTGTACTGTGTTCTCATCCACAGAGGCCAGGATGGGCCAGCTGGAGCC 2714

Db 1865 TTGATGCCAGCTGTACTGTGTTCTCATCCACAGAGGCCAGGATGGGCCAGCTGGAGCC 1924

QY 2715 GTGGCCGAGGGCTTCTCTGATGTGTCTTCTTCTGCTGTGCTTCCGAATCTTT 2774

Db 1925 GTGGCCGAGGGCTTCTCTGATGTGTCTTCTTCTGCTGTGCTTCCGAATCTTT 1984

QY 2775 TAGGCCACTCCAGAGACTCTTAAAGGGGCTTATTCCTGAGAGCCCGAGAGTGGTTG 2834

Db 1985 TAGGCCACTCCAGAGACTCTTAAAGGGGCTTATTCCTGAGAGCCCGAGAGTGGTTG 2044

QY 2835 CCAGTCCCTGCAAGTCCATAGAGTGGAGTGAAGAGATGCTTGTGCATCAATCCCTCAT 2894

Db 2045 CCAGTCCCTGCAAGTCCATAGAGTGGAGTGAAGAGATGCTTGTGCATCAATCCCTCAT 2104

QY 2895 CCAGTCCCTTCT 2954

Db 2105 CCAGTCCCTTCT 2164

QY 2955 CTGCTTCTCTCTATCTCTGCAATTTCTCCGAATCTCTAGGGGATATCCCTATGTCCTCA 3014

Db 2165 CTGCTTCTCTCTATCTCTGCAATTTCTCCGAATCTCTAGGGGATATCCCTATGTCCTCA 2224

QY 3015 GCGCCGTGACTGCCCGACGCTCATTTTTCAGTCAAGTTCGTCCTCTCTCAGGCTTA 3074

Db 2225 GCGCCGTGACTGCCCGACGCTCATTTTTCAGTCAAGTTCGTCCTCTCTCAGGCTTA 2284

QY 3075 TGAAGTCTCAAGTCAAGGACGACCCCTAATCAAGATGGCCAAATCCCTGTGTCTTCC 3134

Db 2285 TGAAGTCTCAAGTCAAGGACGACCCCTAATCAAGATGGCCAAATCCCTGTGTCTTCC 2344

QY 3135 TTGTGTCTTGTCTTATTTGGAGTAGAGTGTCTTACCCCTGTCTGTGGGTGGGTGT 3194

Db 2345 TTGTGTCTTGTCTTATTTGGAGTAGAGTGTCTTACCCCTGTCTGTGGGTGGGTGT 2404

QY 3195 GTTTCAGACAGCTGCTTCTGTGCATTTGTGTCTGTGCTGCTCATGTCTATAGGGA 3254

Db 2404 GTTTCAGACAGCTGCTTCTGTGCATTTGTGTCTGTGCTGCTCATGTCTATAGGGA 3254

Db	2405	GTTTAGAGACAGCTGCTCTGTGCATTTGTGTCTGCTGCTCAGCTCTCATATAGAGA	2464
QY	3255	GGATGCTCATCGTGCACAGCACAGCTCAAGTTAGCATTTCAAGTATTTGGGGGTGCAT	3314
Db	2465	GGATGTCATCGTGCACACACAGCTCAAGTTAGCATTTCAAGTATTTGGGGGTGCAT	2524
QY	3315	GATTAATGAAGAAATGGCCATTTTGTACCAAGGGCTCTGTATTTCTGCAACAGCTGTTGGCA	3374
Db	2525	GATTAATGAAGAAATGGCCATTTTGTACCAAGGGCTCTGTATTTCTGCAACAGCTGTTGGCA	2584
QY	3375	GGCTGAGTGGAAACCAAAAGGCTGGGCATCAAAAGATGAAAGACCAAGCCCTACAACTCC	3434
Db	2585	GGCTGAGTGGAAACCAAAAGGCTGGGCATCAAAAGATGAAAGACCAAGCCCTACAACTCC	2644
QY	3435	AGCCACCCAGCCAGAGAGGGGCTGTCCATTCACATTCAGGCAATGCGAATGAGCTGGCCCT	3494
Db	2645	AGCCACCCAGCCAGAGAGGGGCTGTCCATTCACATTCAGGCAATGCGAATGAGCTGGCCCT	2704
QY	3495	GGGTGAGTGGGGGCTGTGGCTAGTGGGAGAGGGGCTGGCCGGTGGGGGCAAGGGCCCTGG	3554
Db	2705	GGGTGAGTGGGGGCTGTGGCTAGTGGGAGAGGGGCTGGCCGGTGGGGGCAAGGGCCCTGG	2764
QY	3555	CGTGCTCAGGCTTGGGCTCCATTCCTCCATCACTGCTGTCCTCTGAGAGTCTGCAATTTGG	3614
Db	2765	CGTGCTCAGGCTTGGGCTCCATTCCTCCATCACTGCTGTCCTCTGAGAGTCTGCAATTTGG	2824
QY	3615	GATGGGACCAAAAGAAATAGCAAGAGATAGAAACAACAGAAATTTTCTCTTAAAGAC	3674
Db	2825	GATGGGACCAAAAGAAATAGCAAGAGATAGAAACAACAGAAATTTTCTCTTAAAGAC	2884
QY	3675	TGGTTAAATCAATTCTGTATACAGCCTTCACATTAATGTATGCGCTAAAAAATAATTG	3734
Db	2885	TGGTTAAATCAATTCTGTATACAGCCTTCACATTAATGTATGCGCTAAAAAATAATTG	2944
QY	3735	TATGCTTTATATACATAATATGTATAATCTTCAGGTAAAAAAGGCAAGCCACAGAAATG	3794
Db	2945	TATGCTTTATATACATAATATGTATAATCTTCAGGTAAAAAAGGCAAGCCACAGAAATG	3004
QY	3795	TGTTATAGGCACTTCCCATTTGTGTGTATAGAAAGAGTAGAATATAAACACTAATTTGCT	3854
Db	3005	TGTTATAGGCACTTCCCATTTGTGTGTATAGAAAGAGTAGAATATAAACACTAATTTGCT	3064
QY	3855	TATGTATGCTCATTTCCAGAAATGAATGGGTAAACACTGATTACTTTGGAGGGGAAACAGTA	3914
Db	3065	TATGTATGCTCATTTCCAGAAATGAATGGGTAAACACTGATTACTTTGGAGGGGAAACAGTA	3124
QY	3915	GGTGAGGACACGAGAGGAAAGGCTTTAAACACTTACACCCCTTTGTACATTTTGAATTT	3974
Db	3125	GGTGAGGACACGAGAGGAAAGGCTTTAAACACTTACACCCCTTTGTACATTTTGAATTT	3184
QY	3975	TGAACCAATGACGTATTTACCTATTCAAAATTAACAAATTAATGGGGCC 4023	
Db	3185	TGAACCAATGACGTATTTACCTATTCAAAATTAACAAATTAATGGGGCC 3233	

RESULT	10
LOCUS	BTV15774
DEFINITION	Bos taurus mRNA for copper amine oxidase.
ACCESSION	J15774
VERSION	J15774.1
KEYWORDS	GI:2665357
SOURCE	copper amine oxidase. cow.
ORGANISM	Bos taurus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos. 1 (bases 1 to 2605)
AUTHORS	Hogdall,E.V., Houen,G., Borre,M., Bundgaard,J.R., Larsson,L.I. and Vunst,J.
TITLE	Structure and tissue-specific expression of genes encoding bovine copper amine oxidases
JOURNAL	Eur. J. Biochem. 251 (1-2), 320-328 (1998)

Medline	98151264
PubMed	9492300
Reference	2 (bases 1 to 2605)
Authors	Vuust, J.
Title	Direct Submission
Journal	Submitted (03-DEC-1997) J. Vuust, Statens Serum Institut, Artiklerives 5, DS-2300 Copenhagen, DENMARK
Comment	Related sequence: L27218.
Features	Location/Qualifiers
Source	1..2605
gene	/organism="Bos taurus"
	/db_xref="taxon:9913"
	/sex="Female"
	/tissue_type="lung"
	/clone_id="Clontech"
	/dev_stage="adult"
	12..2300
	/gene="BOLA0"
	12..2300
	/gene="BOLA0"
	/EC_number="1.4.3.6"
	/codon_start=1
	/product="copper amine oxidase"
	/protein_id="CAI576.1"
	/db_xref="GI:2665358"
	/db_xref="SPTREMBL:O46406"
	/translation="MFIFIFSLTLLVLMGREGVSGSEGVGKCHSLPPRCPSRS PSODPMTHPDOSOLFADLSREBELVAVSPGLOKPDYLAQARSPDNCTFSVELD PKRAALAHDRSPPPARELAIVFFGSGPQRPVATLVLPGLPSPYMDVIERGG GLPLRYRHVTLREYLDIDQMTFRELPLQAGVLHCCSYKQGGNLTMTAPPGQLQ SGDAWTFGLYINISKAGYLLHPVGLLVDHKLADPOMTIOVFPGGRYSELVOL EEOEGARVWVVIIPNNGTSGWSLKSQVPPGPPIQFPHQGRFVQSGRSVLSLM TFESGLAGSGSPRIATYRFOGERLAYEISLOEVAIYGNTPAAMLRMYDCGMMK PAPTLRGVDCSPYATYVDVHFLLESQAPTLHARCFVFDNGLPLRRHSDPISY FGVYEVYLFRRSYTLTNDYDMYFHPNGALVFPHTGTISAFEFEGTAKIGN QVRNLTGTVHTSHSAHYKVDLDVGLGNWMAEDMARVPTVPWSPHQJQRLQVIRK OLEEEDQAFPLPHTSGSERYLYLASKONSKHPRGYIQTVSPAGRPLNONSSTERA I SWGYSQVLAIVQKRETEPSSSSSVFMONDPMTVPDFAFINNETIAGKDLVAMVAGPL HIIPIAEDIPNTVYGVNGVGFELRPYNFDEDPMSDADSIYFRGDDAGSEINPLAC LPOAKTCAPDLVPVSHGYPEY"
BASE COUNT	543 a 760 c 734 g 568 t
ORIGIN	
Query Match	44.2%; Score 1787.6; DB 4; Length 2605;
Best Local Similarity	82.5%; Pred. No. 0;
Matches 2065; Conservative	0; Mismatches 429; Indels 10; Gaps 1;
QY	221 GTTGTGTCCTGCTGCTGGCGCAGGGGGTGAAGTGGGGGTGAACCCAGCCAGCTTCCCAT 280
Db	69 GGTGCTGTGGGAGTGAAGAGGAGGAGTTGGGAAGCATGTCACTCCAGCCTGCTCCCGC 128
QY	281 TGGCCCTCTGATATCTCCAGTGGCCAGGCTTGAGACACACCTGGCCAGAGCAGCTGTT 340
Db	129 TGGCCCTCTCAAGTATCCCTAGTGAACAGCCCTGGGACACCTTGACCAAGCAGCTGTT 188
QY	341 GCAGACTGAGCCGAGAGAGACTGACGGCTGTGATGACCTTCTTGACCCAGCAGGCTGGG 400
Db	189 GCAGACTGAGCCGAGAGAGACTGACACTGTGAAGAGACTTCTTGACCCAGAGCTGGGG 248
QY	401 CCAGGCTGTGTGATGACAGCCAGGCCCGGCTTCGACAACTGTCTCTCACTGAGG 460
Db	249 CCAGACTGTGTGATGACAGCCAGGCCCGGCTTCGACAACTGCATCTCTCGGTAGAG 308
QY	461 TTGCAAGCTGCTCCCAAGGCTGACAGCCCTGTGCTACTTGTGACAGAGGGGAGCCCCCACT 520
Db	309 CTGCAAGCTGCTCCCAAGGCTGACAGCCCTGTGCTACTTGTGACAGAGGAGGAGCCCCCACT 368
QY	521 GCCCGGAGAGCACTGGCCATGCTCTTCTTGGACAGCAACCCCAAGCCCAAGTGAAGTGA 580
Db	369 GCCCGGAGAGCACTGGCCATGCTCTTCTTGGTGAACAAACCCCAAGTGAAGTGAAG 428
QY	581 CTGTGTGTGGGAGCACTGCTTACCCCTCTTCAATGCGGAGCTGACTGTGGAGCGTAT 640

DEFINITION	Homo sapiens clone 23577 placenta copper monamine oxidase mRNA, partial cds.
ACCESSION	AF054985
VERSION	AF054985.1 GI:3005693
KEYWORDS	
SOURCE	
ORGANISM	Homo sapiens. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	AUTHORS Anderson,B., Wentland,M.A., Ricefrente,J.Y., Liu,W. and Glibbs,R.A. TITLE A double adaptor method for improved shotgun library construction JOURNAL Anal. Biochem. 236 (1), 107-113 (1996) MEDLINE 96207227 PUBMED 8619474
REFERENCE	AUTHORS Yu,W., Anderson,B., Worley,K.C., Muzny,D.M., Ding,Y., Liu,W., Ricefrente,J.Y., Wentland,M.A., Lennon,G. and Glibbs,R.A. TITLE Large-scale concatenation cDNA sequencing JOURNAL Genome Res. 7 (4), 353-358 (1997) MEDLINE 97264341 PUBMED 9110174
REFERENCE	AUTHORS Yu,W. and Glibbs,R.A. TITLE Direct Submission JOURNAL Submitted (23-MAR-1998) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza S930, Houston, TX 77030, USA LOCATION/Qualifiers 1..1794 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE Consortium 23577" /sex="female" /tissue-type="brain" /clone_lib="LIB" /dev_stage="infant" 1..196 /note="similar to Homo sapiens placenta copper monamine oxidase encoded by Genbank Accession Number U93447"
CDS	/codon_start=2 /product="placenta copper monamine oxidase" /protein_id="AAC09346.1" /db_xref="GI:3005694" /translation="GVGFRLRPYNFEDDEFSVSADSIYRGDDAGACEVNPPLACLPPQAAACAPDLPAFISHGGRFSHN" BASE COUNT 397 a 1482 c 455 g 460 t ORIGIN
Query Match	44.0%; Score 1779.2; DB 9; Length 1794; Best Local Similarity 99.8%; Pred. No. 0; Matches 1781; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY	2257 CGGGCTGGGCTTCTCCGCCACCCTATTAACCTTTGTACGAAGACCCCTCTTACTC 2316 Db 1 CCGGCTGGGCTTCTCCGCCACCCTATTAACCTTTGTACGAAGACCCCTCTTACTC 60
OY	2317 TGCCAGTCATCTACTTCCGAGGGAGACAGATGCTGGGCTCGAGATCAACCCCT 2376 Db 61 TGCCAGTCATCTACTTCCGAGGGAGACAGATGCTGGGCTCGAGATCAACCCCT 120
OY	2377 AGCTTGCTGCCCGAGSGCTGCTGTGCCCCGACCTCCCTGCCTTCTCCAGGGGG 2436 Db 121 AGCTTGCTGCCCGAGSGCTGCTGTGCCCCGACCTCCCTGCCTTCTCCAGGGGG 180
OY	2437 GTTCTCTACAACATAGGGGCTCTGGGATGGGGCATGTGGCCAGGGCTCCAGGGCCAGG 2496 Db 181 GTTCTCTACAACATAGGGGCTCTGGGATGGGGCATGTGGCCAGGGCTCCAGGGCCAGG 240
OY	2497 GTGTAGAGGATGGGAGAGCTGGGACATGGGCGCGAGCTGGTTCCTTTCTGTG 2556 Db 241 GTGTAGAGGATGGGAGAGCTGGGACATGGGCGCGAGCTGGTTCCTTTCTGTG 300
OY	2557 CCAGAGCTCTCTTCTTCACATTACCCTCCGTGCATCCGCCCTTAGCCAGAGCTCTCT 2616

Db	301	CCAGGAC	CTCTTCTTCCATCCACCTACCCCTCCGATCCGCGCTGTGAGCCAGGAGCTCT	360
Qy	2617	GACCC	TGTAAGCTTGACACAGGGGACACTGAACCTTTGTATGCGAGCTGTACTGAATT	2616
Db	361	GACCC	TGTAAGCTTGACACAGGGGACACTGAACCTTTGTATGCGAGCTGTACTGAATT	420
Qy	2677	CTCAT	CCACAGAGCCAGGAGTGGCCAGCTGGACCGTGGGCCGAGGGGCTCCCTAGAT	2736
Db	421	CTCAT	CCACAGAGCCAGGAGTGGCCAGCTGGACCGTGGGCCGAGGGGCTCCCTAGAT	480
Qy	2737	GGTTC	CCCTTGTGCTGTGCGCTTTCGCCAATCTTTTAAAGCCACTTCAAGACTGA	2796
Db	481	GGTTC	CCCTTGTGCTGTGCGCTTTCGCCAATCTTTTAAAGCCACTTCAAGACTGA	540
Qy	2797	AAAGG	GGGCTATTCCCTGGAGACCCCGAGTAGGGTTGGCAGCTCGCAAGTCCATAGCT	2856
Db	541	AAAGG	GGGCTATTCCCTGGAGACCCCGAGTAGGGTTGGCAGCTCGCAAGTCCATAGCT	600
Qy	2857	GAGCG	GAAGAGATGCTTGTGCTCACATTCGCCCTATCCAGAGTCCCTTCCCTGCGCT	2916
Db	601	GAGCG	GAAGAGATGCTTGTGCTCACATTCGCCCTATCCAGAGTCCCTTCCCTGCGCT	660
Qy	2917	TCTCT	CTCTACCTACTCTCTCCCTCCCTCCCTGTTTCTTCCGCTTCTCTTATCTGCA	2976
Db	661	TCTCT	CTCTACCTACTCTCTCCCTCCCTCCCTGTTTCTTCCGCTTCTCTTATCTGCA	720
Qy	2977	ATTTC	CCCCGAATCCTAGGGGATATCCATATCCACGCCCTGTACTCCCCAGCC	3036
Db	721	ATTTC	CCCCGAATCCTAGGGGATATCCATATCCACGCCCTGTACTCCCCAGCC	780
Qy	3037	TCAGT	TTCCATCAAGTTCGCTCCCTCCCTCCAGCCCTAAGGAAGTCTCAAGTTCACGGGA	3096
Db	781	TCAGT	TTCCATCAAGTTCGCTCCCTCCCTCCAGCCCTAAGGAAGTCTCAAGTTCACGGGA	840
Qy	3097	CCCCA	TACAGAGGACCAATCCCTGTGTCTCTCCCTGTGTGCTGTATATGGA	3156
Db	841	CCCCA	TACAGAGGACCAATCCCTGTGTCTCTCCCTGTGTGCTGTATATGGA	900
Qy	3157	GTAGA	GTGCTCTCTACCCCTGTCTGTGGGCTGGGTGTGTTTCAGACAGCTGCTTCTGT	3216
Db	901	GTAGA	GTGCTCTCTCTACCCCTGTCTGTGGGCTGGGTGTGTTTCAGACAGCTGCTTCTGT	960
Qy	3217	GCATT	TGTGTGCTGCTGCCTGCATGCTCTATATAGAGAGATAGTGCATGTACAGCAGC	3276
Db	961	GCATT	TGTGTGCTGCTGCCTGCCTGCATGCTCTATATAGAGAGATAGTGCATGTACAGCAGC	1020
Qy	3277	AGCTC	AAGTTAGCATTTCAAGTATTTGGGGGTGCAATGATATGAAGAATGGCCATTTT	3336
Db	1021	AGCTC	AAGTTAGCATTTCAAGTATTTGGGGGTGCAATGATATGAAGAATGGCCATTTT	1080
Qy	3337	GTAGC	AGGGCTCTTATTTCTGCAACAGCCTGTTTGGAGAGCTGTGAGTGGAAACAAGGCT	3396
Db	1081	GTAGC	AGGGCTCTTATTTCTGCAACAGCCTGTTTGGAGAGCTGTGAGTGGAAACAAGGCT	1140
Qy	3397	GAGCA	TCAAATAGAGAAGCCAAAGCCCTCAACAATCCAGCCACCCAGCCAGAGAGGGCT	3456
Db	1141	GAGCA	TCAAATAGAGAAGCCAAAGCCCTCAACAATCCAGCCACCCAGCCAGAGAGGGCT	1200
Qy	3457	GTCCA	TACATTCAGGCATGCGAATGAGCTGGGCCCTGGGTGAGGTGGGGTCTGCGCT	3516
Db	1201	GTCCA	TACATTCAGGCATGCGAATGAGCTGGGCCCTGGGTGAGGTGGGGTCTGCGCT	1260
Qy	3517	AGTGG	AGAGGGGCTGGCTGTGGGTGGGGAGAGGGCTGTGCTGTCCAGGCTTGGGCTCA	3576
Db	1261	AGTGG	AGAGGGGCTGGCTGTGGGTGGGGAGAGGGCTGTGCTGTCCAGGCTTGGGCTCA	1320
Qy	3577	TTCCCA	TACATGCTGTGCCCTGTGAGGTGGAATGGGAGTGGGGAACAAGAAATAGCAA	3636
Db	1321	TTCCCA	TACATGCTGTGCCCTGTGAGGTGGAATGGGAGTGGGGAACAAGAAATAGCAA	1380
Qy	3637	GAGAT	GAGAAACAAGAAATTTTTTCTTAAAGACTGTGTTAATCAATCTGTATCA	3696

Db 1381 GAGATGACAAACAGAACTTTTCTCTTAAGGACTGTTAAATCAATTCGATACA 1440
QY 3697 GCCTTACATACATAGTATGACGCTAAATAATATGTATGTCTTATATCTAATATG 3756
|||||
Db 1441 GCCTTACATACATAGTATGACGCTAAATAATATGTATGTCTTATATCTAATATG 1500
QY 3757 TAAATATCTTACAGTGAAGAAAGGCAAGCCACAGAAATGTATAGCGACTTCCATTGG 3816
|||||
Db 1501 TAAATATCTTACAGTGAAGAAAGGCAAGCCACAGAAATGTATAGCGACTTCCATTGG 1560
QY 3817 TGTTCACAAGGAGTAAATATTAACATATGCTATGTATAGCGCTATTCAGAAATPA 3876
|||||
Db 1561 TGTTCACAAGGAGTAAATATTAACATATGCTATGTATAGCGCTATTCAGAAATPA 1620
QY 3877 ATGGCTAACACTGATTAATCTTTGGGAGGGAACACAGTAGCTGAGAGACAGAGGGAAG 3936
|||||
Db 1621 ATGGCTAACACTGATTAATCTTTGGGAGGGAACACAGTAGCTGAGAGACAGAGGGAAG 1680
QY 3937 GGTCTTAACACTTACACCCCTTTTGTACATTTTGAATTTTGAACCATGTGACTGTATACC 3996
|||||
Db 1681 GGTCTTAACACTTACACCCCTTTTGTACATTTTGAATTTTGAACCATGTGACTGTATACC 1740
QY 3997 TATTCAAAATTAACATTAATATGAGCCCAACAGAAAAA 4040
|||||
Db 1741 TATTCAAAATTAACATTAATATGAGCCCAACAGAAAAA 1784
|||||

RESULT 12
AB050500S1 2614 bp DNA linear PRI 05-DEC-2000
DEFINITION Homo sapiens AOC3 gene for vascular adhesion protein-1, exon 1.
ACCESSION AB050500
VERSION AB050500.1 GI:11602470
KEYWORDS
SEGMENT
SOURCE
1 of 3
Homo sapiens pre-pro-B cell cell_line:FLB14-14 DNA, clone_11b:keio
BAC library clone:KB51F9.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
TITLE
1
Imamura, Y., Kudoh, J., Mashima, Y., Ogunchi, Y. and Shimizu, N.
Tandem array of two human copper amine oxidase genes (AOC2 and
AOC3) in 17q21
JOURNAL
Unpublished
2 (bases 1 to 2614)
REFERENCE
AUTHORS
TITLE
Direct Submission
Submitted (25-OCT-2000) Nobuyoshi Shimizu, Keio University School
of Medicine, Department of Molecular Biology, 35 Shinanomachi,
Shinjuku-ku, Tokyo 160, Japan (E-mail:shimizu@med.keio.ac.jp,
Tel:03-3351-2370, Fax:03-3351-2370)
JOURNAL
Location/Qualifiers
FEATURES
source
1.2614
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17q21"
/clone="KB51F9"
/cell_line="FLB14-14"
/cell_type="pre-pro-B cell"
/clone_11b="keio BAC library"
432..437
/note="SP-1 site"
453..458
/note="SP-1 site"
479..2238
/gene="AOC3"
479..2238
/gene="AOC3"
/product="vascular adhesion protein-1"
/note="CDS is reported in Acc#AB050502.
homolog of semicarbazide-sensitive amine oxidase

placental monoamine oxidase"
/number=1
/EC_number="1.4.3.6"
2239..2614
/gene="AOC3"
BASE COUNT 516 a 757 c 732 g 609 t
ORIGIN
Query Match 43.6%; Score 1761; DB 9; Length 2614;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1761; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCTTCCACCCCTTAGTCCAGGATCGATACCGGAACTCAGCCAGAGTCCGGA 60
|||||
Db 479 GTCTTCCACCCCTTAGTCCAGGATCGATACCGGAACTCAGCCAGAGTCCGGA 538
61 GCCCCCAACCCCGTCCAGAGGCCAACAGAGAGCCCGCTTGTGCGGTGAGAAATATG 120
|||||
Db 539 GCCCCCAACCCCGTCCAGAGGCCAACAGAGAGCCCGCTTGTGCGGTGAGAAATATG 598
121 CTCTCTTGTGTAATCAGCTGTCCCTTGTGTGGAAATGAACCAAGACAATCT 180
|||||
Db 599 CTCTCTTGTGTAATCAGCTGTCCCTTGTGTGGAAATGAACCAAGACAATCT 658
181 GTCTCTCTCATTTTGGCGGTATCATCATCTTTTGGCTTGTGTCTCTGTGTGG 240
|||||
Db 659 GTCTCTCTCATTTTGGCGGTATCATCATCTTTTGGCTTGTGTCTCTGTGTGG 718
241 CAGGGGTGAGATGGGGGTGAACCCAGCAGAGTCCCATTCGCCCTGTATCTCCAG 300
|||||
Db 719 CAGGGGTGAGATGGGGGTGAACCCAGCAGAGTCCCATTCGCCCTGTATCTCCAG 778
301 TCCCCAGCCTTGGAGACACCTCTGGCCAGAGCCAGCTGTTTGCAGACTGAGCCGAGAGA 360
779 TCCCCAGCCTTGGAGACACCTCTGGCCAGAGCCAGCTGTTTGCAGACTGAGCCGAGAGA 838
361 GTGACGGCTGTGATGGCTTCTGACCCAGAGGCGTGGGGCCAGGGCTGTGTGATCAG 420
|||||
Db 839 GTGACGGCTGTGATGGCTTCTGACCCAGAGGCGTGGGGCCAGGGCTGTGTGATCAG 898
421 CCAGGCCCGCCCTCGAGACACTGTCTTCTCAGTGGAGTTGACACTGCTCCCAAGC 480
|||||
Db 899 CCAGGCCCGCCCTCGAGACACTGTCTTCTCAGTGGAGTTGACACTGCTCCCAAGC 958
481 TGCAGCCCTGTGCTCACTTGGAGACAGGGGGAGCCCCCACTGCCCGGGAGGCACTGGCCAT 540
959 TGCAGCCCTGTGCTCACTTGGAGACAGGGGGAGCCCCCACTGCCCGGGAGGCACTGGCCAT 1018
541 CTCTCTTCTTGGAGGACCAACCCAGCCCAACGATGAGCTGTGTGGGGCACTGGC 600
|||||
Db 1019 CTCTCTTCTTGGAGGACCAACCCAGCCCAACGATGAGCTGTGTGGGGCACTGGC 1078
601 TCACCCCTCTACATCGGGAGCTGACTGTGAGAGCTCATGAGAGGCCCTTGCCTATTA 660
|||||
Db 1079 TCACCCCTCTACATCGGGAGCTGACTGTGAGAGCTCATGAGAGGCCCTTGCCTATTA 1138
661 CCAGGCCCGCCCTGTCTTCCAGAGTACTGTGACATAGACAGATGATCTTCAACAGAGA 720
1139 CCAGGCCCGCCCTGTCTTCCAGAGTACTGTGACATAGACAGATGATCTTCAACAGAGA 1198
721 GTGCCCCAGGCTTGTGGCTTCTCAACACTGTGCTTCAACAGAGCAGGGGAGGGA 780
|||||
Db 1199 GTGCCCCAGGCTTGTGGCTTCTCAACACTGTGCTTCAACAGAGCAGGGGAGGGA 1258
781 CTTGTGACATAGACAGGCTCCCGTGTGCAATGAGGGAGCCGACCACTGTT 840
|||||
Db 1259 CTTGTGACATAGACAGGCTCCCGTGTGCAATGAGGGAGCCGACCACTGTT 1318
841 TGGCTCTTACTACAACTTCGGGCGCTGGGTTCTTCTCTCACACAGCTGGGCTTGAAGCT 900
|||||
Db 1319 TGGCTCTTACTACAACTTCGGGCGCTGGGTTCTTCTCTCACACAGCTGGGCTTGAAGCT 1378
901 GCTAGTGACCAAGAGCCCTTGACCTGCCCGCTGACTATCCAGAGAGGTGTTCTATCA 960

```

Db 1379 GCTAGTACACACAGGCGCTTGACCTGGCCGCTGACTATCCAGAGGCTGCTATCA 1438
Qy 961 AGCGCGCTACTACAGACAGCCTGGCCAGCTGGAGAGCCAGTTTGAAGCCGCGCTGTGAA 1020
Db 1439 AGCGCGCTACTACAGACAGCCTGGCCAGCTGGAGAGCCAGTTTGAAGCCGCGCTGTGAA 1438
Qy 1021 TGTGCTGCTATCCACACATAGGACAGTGGGTCTGCTGCTGCTGAAGTCCCTGTGCC 1080
Db 1499 TGTGCTGCTATCCACACATAGGACAGTGGGTCTGCTGCTGCTGAAGTCCCTGTGCC 1558
Qy 1081 CCCGGGTACAGCTCCCTCTACAGTCTATCCCAAGGCGCCGCTCAGTGTCCAGG 1140
Db 1559 CCCGGGTACAGCTCCCTCTACAGTCTATCCCAAGGCGCCGCTCAGTGTCCAGG 1618
Qy 1141 AAGTCAGTGGCCCTCTACAGTGTGAGATTTCTCCCTTGGCCCTGAGACATTCAGTGGCC 1200
Db 1619 AAGTCAGTGGCCCTCTACAGTGTGAGATTTCTCCCTTGGCCCTGAGACATTCAGTGGCC 1678
Qy 1201 AAGGATCTTTGACCTTGGCTTCCAGAGAAAGACTGTTTATAGATTAAGCTTCCAGA 1260
Db 1679 AAGGATCTTTGACCTTGGCTTCCAGAGAAAGACTGTTTATAGATTAAGCTTCCAGA 1738
Qy 1261 GACCTTGCCATCTATGTTGGAATCCCGACAGCAATGACACCGCTATGTTGATGG 1320
Db 1739 GACCTTGCCATCTATGTTGGAATTTCCCGACAGCAATGACACCGCTATGTTGATGG 1798
Qy 1321 AGGCTTGGCATGGGCAAGTACACAGCCCGCTGAGCGGTGGTGGACTGCTCCCTACTT 1380
Db 1799 AGGCTTGGCATGGGCAAGTACACAGCCCGCTGAGCGGTGGTGGACTGCTCCCTACTT 1858
Qy 1381 GGGCAGCTAGTGGACAGGCACTCTTTTGGATCCCGAGGCGCCCAAGACATACGTGA 1440
Db 1859 GGGCAGCTAGTGGACAGGCACTCTTTTGGATCCCGAGGCGCCCAAGACATACGTGA 1918
Qy 1441 TGCCTTTGCTGTTTGAACAGAACAGGCGCTCCCGTGGGAGACACACTAGATCT 1500
Db 1919 TGCCTTTGCTGTTTGAACAGAACAGGCGCTCCCGTGGGAGACACACTAGATCT 1978
Qy 1501 CTACTGCCACTACTTTGGGGGTCTTGGGAAACGGTGTCTCTCAGATCTATGTCAC 1560
Db 1979 CTACTGCCACTACTTTGGGGGTCTTGGGAAACGGTGTCTCTCAGATCTATGTCAC 2038
Qy 1561 CTTCCTCACTATGATGTTGGGATAGGCTTCCACACCCAGTGGGGCATAGAAAT 1620
Db 2039 CTTCCTCACTATGATGTTGGGATAGGCTTCCACACCCAGTGGGGCATAGAAAT 2098
Qy 1621 ACATCTATGACCAAGGCTACATCAGCTGGCATTCCTTTTGGTCTACTGGAAGTA 1680
Db 2099 ACATCTATGACCAAGGCTACATCAGCTGGCATTCCTTTTGGTCTACTGGAAGTA 2158
Qy 1681 CGGAAACCAAGTGCACAGCACACCTGGGACGGTCCACACCCACAGCGCCACTTCAA 1740
Db 2159 CGGAAACCAAGTGCACAGCACACCTGGGACGGTCCACACCCACAGCGCCACTTCAA 2218
Qy 1741 GGTGATCTGGATGTAGCAGG 1761
Db 2219 GGTGATCTGGATGTAGCAGG 2239

```

RESULT 13

BOVRA
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BOVRA 2664 bp mRNA linear MAM 03-AUG-1994
Bos taurus serum amine oxidase mRNA, complete cds.
L27218.1 GI:520784
amine oxidase.
Bos taurus female liver, bovine cdna to mRNA.
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bovinae; Bos.
REFERENCE
1 (bases 1 to 2664)

```

AUTHORS      Nu,D., Medzhiradzky,K.F., Adams,G.W., Mayer,P., Hines,W.M.,
              Burlingame,A.L., Smith,A.J., Cai,D., and Kliman,J.P.
TITLE        Primary Structures for Mammalian Intracellular and Serum Copper
              Amine Oxidases
JOURNAL      J. Biol. Chem. (1994) In press
FEATURES
Source       location/Qualifiers
            1..2664
            /organism="Bos taurus"
            /db_xref="taxon:9913"
            /sex="female"
            /tissue_type="liver, bovine"
            61..2349
            /codon_start=1
            /product="serum amine oxidase"
            /protein_id="AAA30525.1"
            /db_xref="GI:520785"
            /translation="MFIFLSIMTLVLMRGREGVSGREGVKOCHPSLPKPCRS
            PSQPTHTHDSOLFADLSRELLTMTASTLQDGLDVLDAOARSDKCVSEVLQ
            PKVAALAHLDGSPPPAREALAIIVFGGQPOPNVETLVVGPDPQSYRVDYVERHG
            GPLPYRRPVLNREYLDIDOMFNRELPAAGVLAHHCSSYKOGOKLTMNSAPRGVQ
            SCDRSTWFGIYNYITKGGPYLHPVGLLELDVPHKALDPADMTVOKVFQGRYENLAL
            EOEFGAGVNVVYIPDDGTGFWSLKSOVPGPTPILOHPOGPRPSVGNRYASLSLM
            TSEFGAFSGSPRPVDFVROGERTLAEISLOEGAVYGGNTPAAMLTFRYMSDFGMY
            FAPLIRGDCPYLATYMDHPVYESQTKTLDAICVDEQNKGLPLRRHSPFLSH
            FGVATVLFKRSVSTMLNIDYMDVFTFNGAIEVLAIGTITSAFLFGARRIGN
            QVEHTLGPVHTSAHYKVDLVGGLNNVMAEDMAFPTALPWPSEHOIORLVTRK
            QLETEBOAFPLIGASPRYLYLASKOSNMGHREGRYRIOTVSPAGPMQNSPMEAF
            SMGRYOLAIORKEETEPSSSVFNQNDPPTVDFEFINETIACKDLVAVTTAGFL
            HIPHARDIPMTVTVGVGVGFELRPYNEFPQEPSMDSADSIYFREGDASCEINPLAC
            LPOAATCADLPIFSGGYPEY"
BASE COUNT   549 a 781 c 751 g 583 t
ORIGIN
Query Match  43.4%; Score 1754; DB 4; Length 2664;
Best Local Similarity 81.6%; Pred. No. 0;
Matches 2044; Conservative 0; Mismatches 450; Indels 10; Gaps 1;

```

Qy 221 GTTGTGCTCTGCTGCTGGGAGGCTGGAGATGGGGGTGAACCCAGCTTCCCAT 280

Db 118 GTTGTGCTGGAGTGGAGGAGGAGTGGGAAGCATGTATCCACCTGCTCCCGC 177

Qy 281 TGCCCTCTGTATCTCCAGTGGCCAGCTTGGACACACCTGGCCAGGACGAGCTTT 340

Db 178 TGCCCTCTCCAGATCTCCCTATGTAGCACACCTGTGACACACCTGACACCTTTT 237

Qy 341 GCAGACCTGAGCCGAGAGAGAGCTGATGCTGCTTTCAGACCCAGCGCTGGG 400

Db 238 GCAGACCTGAGCCGAGAGAGAGCTGATGATGATCTTCTGACTCAGACGCTGGG 297

Qy 401 CCAGGCTGTGTGATGCAAGCCAGGCGCCGCTCGGACAACTGTCTCTCAGTGGAG 460

Db 298 CCAGACCTGTGTGATGCAAGCCAGGCGCCGACCTCAGACAACTGTCTCTCGTGA 357

Qy 461 TTGAGCTGCTCCCAAGGCTGACAGCCCTGAGCTACTTGGACAGGGGAGACCCCGCCT 520

Db 358 CTTCAGCTGCGCCCAAGGCTGACAGCCCTGAGCCACTGTGACAGAGGAGACCCCGCCT 417

Qy 521 GCCCGGAGGACATGGCCATGCTTCTTGGCAGGCAACCCCAAGCCCAAGTGAAG 580

Db 418 GCCCGGAGGACATGGCCATGCTTCTTGGCAGCAACCCCAAGCCCAAGTGAAG 477

Qy 581 CTGTGTGTGGGGCCACATGCTCACCCTCTACTATGGGGGACGTGATGGAGCGTCAT 640

Db 478 CTGTGTGTGGGGCCGCGCCCAAGCCCTCTACTATGGGGGATGACAGCGTGAAGCGTCAT 537

Qy 641 GGAAGCCCTGCGCTATCAACGAGCGCCGCTGCTTCCAAAGATACGTGACATAGAC 700

Db 538 GCGCGCCCTGCGCTATTCACAGCGCCCGCTGCTTCCGAGAGTACGTGACATAGAC 597

Qy 701 CAGATGATCTTCAACAGAGAGTGGCCCAAGGCTTGGGCTTCTCCACCACTGTGCTTC 760

Db 598 CAGATGATCTTCAACAGAGAGTGGCCCAAGGCTGCTGCTCTGACACCACTGTGCTTC 657

QY	761	TAAAGACCGGGGAGGAGAACCGGGGAGAAAGACACAGGCTCCCGGGGCTCAATCA	820
Db	658	TACAAACAGAGGAGACAGAAACTGTGGACCATGAACTACAGTCCCGGTGAGTGCAC	717
QY	821	GGGAGCCGGGACACGTGGTTGGCCCTCTACTACAAACATCTCGGGGCGTGGGTTTCTC	880
Db	718	GGTGATAGTGCACCTTGGTTTTGGACATCTATAACATCACAAAGGGGGGCTTACTCTG	777
QY	881	CACCACGTGGGCTTTGAGCTGCTAGTGAACCAAGGCCCTTGACCTTGCCCGCTGACT	940
Db	778	CACCCCTGGGGTTGAGGCTCTGTGGTAGCCATAAGGCGCTGAGACCCCTGCATTTGACC	837
QY	941	ATCCAGAAAGTGTTCATATCAAGGCGGCTCTACAGACAGCGGCGGACCTGGAGGCCAG	1000
Db	838	GTCACAGAAAGTGTTCATCAAGGCGGCTCTACTATGAAAAATCTGGCCCACTGGAGACAG	897
QY	1001	TTTGAGGACCGGCTGGTGATGTGGTGTGATCCAGACAAATGGACAGTGGGTCTCG	1060
Db	898	TTTGAGGCTGGCAGGAGTAATGTGGTGTGATCCAGAAAGTGGACAGTGGGTCTCG	957
QY	1061	TCCTGAAGTCCCTGTGCCCCGGGTCCAAGTCCCCCTTACAGTCTATCCCAAAGC	1120
Db	958	TCCCTGAAGTCCAGAGGCGCTCCGGGTCCAACTCCCGCTTCAGATTCATCTCAAGGC	1017
QY	1121	CCCGGCTTCAGTGTCCAGGGGAATCGAAGTGGGCTCTCAGTGGACATTTCTCTTGGC	1180
Db	1018	CCCGGCTTCAGTGTCCAGGGCAATCGAAGTGGGCTCTCTCATTTGGACATTTCTCTTGGC	1077
QY	1181	CTCGGACATTCAGTGTGCCCCCAAGATCTTTGACGTTCGCTTCCAAGGAAAGACTAGT	1240
Db	1078	CTCGGACATTCAGTGTGCTCTAGGGCTTTGAGCTTCATTCAGAGGAAGAGCGCTGCT	1137
QY	1241	TATGAGATTAAGCTCCAGAAGGCGCTTGGCCATCTATGTGTGAATTTCCCAAGACAATG	1300
Db	1138	TATGAGATTAAGCTCCAGAAGGCGCTGTGCTGTCTACGGTGGAAATCTCAGACAGCAATG	1197
QY	1301	ACGAGCCGATGTGATGGATGGAGGCTTTGGCAATGGGCAATACAGAGCCCGCTGACCGCT	1360
Db	1198	CTCACTGCTATATGATGATGTGGCTTTGGATGGGTACTTGTGCACACCCCTGATTTGCT	1257
QY	1361	GGGGTGGACATCCCTCACTTGTGGCCACCTACAGTGGATGGCACTTCTTTTGGAGTCCAG	1420
Db	1258	GGGCTAGACTGCCCTTACCTGCGGCACCTACATAGACTGGCACTTGCTGTGGAAGTCCAA	1317
QY	1421	GCCCCAAGACATACGTGATGCCCTTTGTGTGTGAAACAGAACACAGGCGCTCCCGCTG	1480
Db	1318	ACCCCAAGACATACATGATGCGCTTTGTGTGTGACAGACAAAGGCGCTCGCCCTG	1377
QY	1481	CGGGACACCACTAGATCTCTACTGGCACTATTGGGGGCTTTCGGAAACGATGCTG	1540
Db	1378	AGGGACACCACTAGATTTTCTTCCACTATTTTGGGGGGGCTTGCACAGACATGTCTG	1437
QY	1541	GTGCGACAGATTAATGTCCACTTGTGCACATAGATATGTGGGATATACGCTCTTCCAC	1600
Db	1438	GTCTTCAGGTGTGTGCACAGATGTCACTAATAGATATGTGGGATATGTGCTTCTTAC	1497
QY	1601	CCCAATGGGGCCATAGAAATACGATCTTATATGCCACAGGCTACATGACATCGGACTTCTC	1660
Db	1498	CCCAATGGGGCCATAGAAAGTCAATTGGCAATGCCAGGGCTACATGACGTCAACGTTCTC	1557
QY	1661	TTTGTGCTACTGGGAAGTACGGGAACCAAGTCTACAGACACACCTGTGGCACGCTCCAC	1720
Db	1558	TTTGTGCTGCCCAAAATATACGGAACACAGGTTGGGGAACACACGCTGTGGCCCTCTCAC	1617
QY	1721	ACCCACACGGGCCACTCAAGGTGGATCTGGATGTAGCAAGCACTGGAAACATGGGGTCTGG	1780
Db	1618	ACCCACATGCGCACTCAAAAGTGGATCTGAGCTGGAGAGCTGGAAAATCTGGGTCTGG	1677
QY	1781	GCCGAGGATATGCTTTTGTCCCATAGGCTGCTCCCTGAGAGCCCTACAGCAGCTGAG	1840
Db	1678	GCTGAGGATATGGGCTTTTGTGCCACAGGGGCAATACCTGAGAGCCCTGAGCAACAATATAG	1737
QY	1841	AGGCTGACAGTGAACCCGGAAGCTCTGAGATGAGAGACAGAGCGGCTTCTCTGTGGGA	1900

Db	1738	AGGCTGCAAGGTGACCCGGAGAGAGTGGAGAACGAGGACGAGGCTTCCTTCCCTCCCTGGGA	1979
Qy	1901	AGCGCCACCCTCGCTACCTGTACTCTGGCCAGACCAACACAGACACAACTGGGGTACCCC	1960
Db	1798	GGGGGCTTCCCTCCCTACCTGTACTCTGGCCAGACCAAGCAGAGACACAACTGGGGGACCCCT	1857
Qy	1961	CGGGGCTACCGCATCCAGATGCTCAGCTTTGGTGGAGAGCCGGCTCCCAAAAGCTCC	2020
Db	1858	CCCGGCTATTCGATCCAGACGCTCAGCTTTGGTGGGGGGCGGATCCCCAGAACGCCCC	1917
Qy	2021	ATGCGAGAGAGCGTTACGCTGGAGAGAGTACACAGCTGGCTGTGACCCAGCCGAGAGAGAG	2080
Db	1918	ATGGAGAGAGAGCGTTACGCTGGGGGAGGTACACAGCTGGCGATTAACCCAGAGAGAGACA	1977
Qy	2081	GAGCCCAAGTACGAGCAGCGTTTTCATCAGAAATGACCCCTTGGGGCCCACTGTGGATTTC	2140
Db	1978	GAGCCCAAGTACGAGCAGCGTTTTCATCAGAAATGACCCCTTGGAGCTCCACGCTGACTTT	2037
Qy	2141	AGTGAATTCATCAATCAATGAGACCAATGCTGGAAAGAGATTGGTGGCTGGGTGACACT	2200
Db	2038	TCTGATTCATCAATCAATGAGACCAATGCTGGAAAGAGATTGGTGGCTGGGTGACACC	2097
Qy	2201	GGTTTCTGCATATCCACATGACATGACATGACATGACATGACATGACATGACATGACATGAC	2260
Db	2098	GGTTTCTGCATATCCACATGACATGACATGATATCCCAACAGGCTACGCTGGGGAACGCT	2157
Qy	2261	GTCGGCTTCCTCCGACCCCTATACCTCTTTGACGAAAGACCCCTCTTACTGTGCC	2320
Db	2158	GTCGGCTTCCTTGTGAGACCCCTACAACTCTTTGACAGAGACCCCTCCATGATTCGCT	2217
Qy	2321	GACTCATATCTACTCCGAGAGGACCAAGATGCTGGGGCTCGGAGGTCAACCCCTACTGT	2380
Db	2218	GACTCATATCTACTCCGAGAGGACCAAGATGCTGGGTCTGTGAATCAACCCCTGCT	2277
Qy	2381	TGCGTCCCGCAGGCTGCTGCTGCGCCCGACCTCCCTGCTCCACGAGGGCTTC	2440
Db	2278	TGCGTCCCGCAGGCTGCTGCTGCGCCCGACCTCCCTGCTTCCTCCAGGGGGCTAC	2337
Qy	2441	TCTCACACTAGGCGGTCTCTGGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT	2500
Db	2338	CCTGATATCTAGTGTGCTCTTGGGGGAGAGATGTGTCTAAGACCCAAAGACAGGAGT	2397
Qy	2501	GAGGATGAGGAGACACTGCGGACACTGCGGCGGACGCTGCTCTCTCTCTCTCTCTCTCTCT	2560
Db	2398	GTCGAGAGGAGGAGGCTGCGGACACTGAG-----TTTCTCCAGCTCCACCCGAG	2447
Qy	2561	GACTCTCTTCTTCCACTACCTCCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2620
Db	2448	GTTCT	2507
Qy	2621	CTGTGATGCTGACACAGGAGGACATGAACTGTTGATGATGACAGCTGTAGTATCTCTCA	2680
Db	2508	CTGTGATGCTGACACAGGAGGACATGAACTGTTGATGATGACAGCTGTAGTATCTCTCA	2567
Qy	2681	TCCACAGAGCCAGGATGCGCCAGGCTGGAAGCCGTGCGGAGG	2724
Db	2568	TCCACAG	2611
RESULT 14			
LOCUS	569583	2664 bp	mRNA
DEFINITION	copper amine oxidase-amloride-binding protein homolog (cattle, liver, mRNA, 2664 nt).		
ACCESSION	569583		
VERSION	569583.1	GI:546215	
KEYWORDS			
SOURCE	Bos taurus liver.		
ORGANISM	Bos taurus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.		

OY	1721	ACCACAGCCGCCACTTCAAGGTGATCTGGATCTACACGACTGGAGAATCGGTCTGG	1780
Db	1618	ACCACAGTGCCCACTACAAGGTGATCTGGACTGGAGAACTGGGAATCTGGTTGG	1677
OY	1781	GCCCAGATATATGGCTTTGTCCCCATGGCTGTGCCTTGGAGCCCTGAGCACACAGCTCAG	1840
Db	1678	GCTAAGATATATGGCTTTGTTCCTCCACGGCCGATACCTCTGGAGCCCTGAGCACACAGATACAG	1737
OY	1841	AGGCTGCAGGTGACCCGGAAAGCTGTGAGATGGAGAGGACAGGCCGCTTCTCTGTGGGA	1900
Db	1738	AGGCTGCAGGTGACCCGGAAAGCTGTGAGATGGAGAGGACAGGCCGCTTCTCTGTGGGA	1797
OY	1901	AGGCCACACCCCTCGCTACCTGTACCTGTGGCCAGCAACCAAGTAGTGGGTACCCC	1960
Db	1798	GGGGCTTCCCTCGCTACCTGTACCTGTGGCCAGCAACCAAGTAGTGGGTACCCC	1857
OY	1961	CGGGGCTAACCGCATCCAGATGCTCAGGCTTTGCTGGAGAGCGCGGCCCAAAGCTCC	2020
Db	1858	C GG G C T T A C C G C A T C C A G A T G C T C A G G C T T T G C T G G A G A G C G G C C C A A A G C T C C	1917
OY	2021	ATGGCAGAGAGGCTTCAGCTGGGAGAGGTACAGCTGGCTGTGACCCAGCGGAAGAGAG	2080
Db	1918	ATGGAGAGAGGCTTCAGCTGGGAGAGGTACAGCTGGCTGTGACCCAGAGGAAGAGACA	1977
OY	2081	GAGCCCACTAGCAGCAGCGTTTTCAATCAGATAACCTTGGGCCCCCATCTGTGGATTTC	2140
Db	1978	GAGCCCACTAGCAGCAGCGTTTTCAATCAGATAACCTTGGGCCCCCATCTGTGGATTTC	2037
OY	2141	AAGTACTTCATCAACAATGAGAACCATTCCTGGAAAAGATTGGTGGCCTGGGTACAGCT	2200
Db	2038	TCTGACTTCATCAACAATGAGAACCATTCCTGGAAAAGATTGGTGGCCTGGGTACAGCC	2097
OY	2201	GATTTCATCATATCCACATGACAGAGACATCTCTTAACACAGTACTGTGGGAACGGC	2260
Db	2098	GATTTCATCATATCCACATGACAGAGATATCTCCAACACAGGTGTGGGAACGGT	2157
OY	2261	GTGGGCTTCTCTCCGACCCCTAATACTTTTGACGAAAGCCCTCTTACTCTGCC	2320
Db	2158	GTGGGCTTCTCTTGTAGACCCCTACAATCTTTTGACGAGAGCCCTCATGATCTGCT	2217
OY	2321	GACCTCATCTACTTCCGAGGGGAGCAGGATGCTGGGGGCTTGGAGGTCAACCCCTAGCT	2380
Db	2218	GACCTCATCTACTTCCGAGGGGAGCAGGATGCTGGGGCTTGTGAGATCAACCCCTAGCT	2277
OY	2381	TGCCTGCCCCAAGGCTGTGCTGTGCCCCCGACCTCCCTGCTTCTCCACAGGGGGCTTC	2440
Db	2278	TGCCTGCCCCAAGGCTGTGCTGTGCCCCCGACCTCCCTGCTTCTCCACAGGGGGCTAC	2337
OY	2441	TCTCACAACTAGGCGGTCTCTGGGATGGGGCATGTGGCCAAAGGGCTCCAGGGCCAGGGTGT	2500
Db	2338	CCTGATGACTAGGTGTGTCTTGGGGCGAGGATGTGTAGGACCCAAAGCAAGAGGAGT	2397
OY	2501	GAGGGATGGGGAGCAGCTGGGACATGGGGCCGGGACCCGTGTCTTCTTCTGTGGCAG	2560
Db	2398	GTGGGAGGGGAGGGGGCTGGGGCACTGAG-----TTTCTTCCACGCTCCACCCAG	2447
OY	2561	GACTCTCTTTCTTCTCACTACCTCCCTGGCATCCGCTCTGTGAGCCAGGAGCCCTGTGAC	2620
Db	2448	GTTCCTCCCTCTCCCATCTTCTGCCCCCTTGGCTCTCCMAACCGCTGTGGAAATTCATAGC	2507
OY	2621	CTGTGATGCTTGACACAGGGGACACTGAACCTTGTGATGCCAGCTGTAGTGAATTCTCA	2680
Db	2508	CTGTGCTGCTTGACACCCATGGGCTTCTCAACCTCTGTGACTCTGTGACCTGTGATCTCTCA	2567
OY	2681	TCCACAGAGGCCCCAGGCTGGCCAGCTTGGAGACCTGGGCCGAGG	2724
Db	2568	TCCACAGAGAAAAAGAAATGGCCAACTGGGAGTCTAAGATGATG	2611
LOCUS	AF115411	4210 bp	mRNA linear ROD 12-OCT-1999

DEFINITION	Mus musculus semicarbazide-sensitive amine oxidase mRNA, complete cds.
ACCESSION	AF115411
VERSION	AF115411.1
KEYWORDS	GI:4185816
SOURCE	Mus musculus.
ORGANISM	Mus musculus.
REFERENCE	Eukaryotes: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae; Mus.
AUTHORS	1 (bases 1 to 4210)
TITLE	Moldes,M., Fève,B. and Pairault,J.
JOURNAL	Molecular cloning of a major mRNA species in murine 3T3 adipocyte lineage: differentiation-dependent expression, regulation, and identification as semicarbazide-sensitive amine oxidase
MDLINE	J. Biol. Chem. 274 (14), 9515-9523 (1999)
PUBMED	99194802
REFERENCE	2 (bases 1 to 4210)
AUTHORS	Moldes,M., Fève,B. and Pairault,J.
TITLE	Direct Submission
JOURNAL	Submitted (21-DEC-1998) Institut Biomedical des Cordeliers, UPRESA 7079 UPMC CNRS, 15,rue de l'Ecole de Medecine, 75270 Paris Cedex 06, France
FEATURES	Location/Qualifiers
source	1..4210
	/organism="Mus musculus"
	/db_xref="taxon:10090"
	/cell_line="3T3-L1"
	166..2463
	/EC_number="1.4.3.6"
	/function="catalyzes the oxidation of primary amines such as benzylamine, methyl amine, N1-acetyl-putrescine, beta phenylethylamine, tyramine and histamine
	/note="transmembrane enzyme; strongly inhibited by semicarbazide, aminoguanidine and hydrazine derivatives such as phenelzine and hydrazine; abundantly expressed in white and brown adipose tissues and the aorta; expressed at a low level in lung and muscle; MSSAO"
	/codon_start=1
	/product="semicarbazide-sensitive amine oxidase"
	/protein_id="AAD09199.1"
	/db_xref="GI:4185817"
	/translation="MTOKTTLVALAVITIFALVCVLLAGSGDGGSLSPDHCPSP
	LPSQPRHPESQSPFADLSPEELTAVNSPTFKHGKPSLVDAQAQRPDNCVSEELD
	LPAAALAHLDKRGPPRPVRRALATITFGGQKRPVSELVVGPRHPMDYIER
	GLPAPYRRPVLDRYODIEMFIHRELPQASGLAHCCFKHQONLMTTARGL
	OSGRATVFGLYNLSGAGFPHRIGLLIDKMLDPAFLTIOKVFQGRYESLT
	LEDFEAGLVVNVLPVNNGTGSGMSKSVPPKAPLPALPHOGRFVQSGVSSSI
	MAFSPGAGFSGPRFEDIROGERVAYEISYOEALTAIGNSPAMSCVYDGSIGI
	KSPPLIGACGCPRIATYVDNHFLESAPRTLRAPKVPFQDNCPLRRHSDPYS
	YKGVGTVGLVRSVSTLNTLDYLDWNVFHPGALIEVFHATGITSAPFFGAGKRP
	NRVAGHTIATVTHSAHKRYLDVAGLNNMAAEDEAVPTIVPQPYOMORLOVTT
	KLLEEEEAAPPLGATRRKYLDAVLSNHSNKGHRGRIQLISFGKPLPESPLEK
	FTWGVYHLAVTORKEEPSSSSIFENQNDPMPTVFTPEISMETIAGDILAVMTY
	LHIAHEIDPNTVTRAGNSGFEELRPYNFDEDPSPHSADSIYFREGDATACEVPL
	CLSLQATCAPEIPASHSQGFARND"
misc_feature	166..180
	/note="intracellular domain; unclassified site"
misc_feature	181..243
	/note="transmembrane-region site"
misc_feature	244..2460
	/note="extracellular domain; unclassified site"
misc_feature	1576..1578
	/note="topaquinone cofactor binding site; unclassified site"
BASE COUNT	940 a 1191 c 1039 g 1040 t
ORIGIN	
Query Match	42.2%; Score 1705.6; DB 10; Length 4210;
Best Local Similarity	84.1%; Pred. No. 0;
Matches 1924; Conservative	0; Mismatches 364; Indels 0; Gaps 0

Db	165	AATGACCCAGAAAGACACCCTTAAGTGTCTCTGCTCTGCTGTATCAATCCATCTTTCCTT	224
Qy	220	GGTTTGTCCTCCGTGCTGGGACGGGTGGAGATGGGGGTGAACCCAGCCAGCTTCCCA	279
Db	225	GGTTTGTCCTCTGCTAGCTGGCAGAGAGAGATGGGGGTGGAGTACGCCAACCTTTC	284
Qy	280	TTGCCCCCTCTGTATGTCTCCCAATGCCAAGCTTGGACACACCCTTGCCAGAGCCAGCTGT	339
Db	285	TTGGCCCCCTCTGTCTCTTAAGTGTCCAGCCCCGGGACACACCCTTAGCCAGAGCCAGCTT	344
Qy	340	TTGCAGACCTGAGACCCAGAGAGAGCTGAGAGGGGTGGATGCGGTTTCTACCCAGCGCTGG	399
Db	345	TTGCAGACCTGAGACCCAGAGAGAGCTGAGAGGGGTGGATGCGGTTTCTACCCAGCGCTGG	404
Qy	400	GCCAGGGCTGGTGGATGTCAGGCCCCAGGCCCGGCCCTCGAGACAATGTGTCTTCACTGGA	459
Db	405	GCCAGGGCTGGTGGATGTCAGGCCCCAGGCCCGGCCCTCGAGACAATGTGTCTTCACTGGA	464
Qy	460	GTTTGACGTGCTCCCTCCCAAGCTGCGAGCCCTGTGCTCACTTGGACAGGGGGAGCCCCACC	519
Db	465	GTTTGACGTGCTCCCTCCCAAGCTGCGAGCCCTGTGCTCACTTGGACAGGGGGAGCCCCACC	524
Qy	520	TTGCCGGAGAGGACATGGGCATATGTCCTTGTGGAGGCAACCCCAAGCTGAGTGA	579
Db	525	TTGTGAGAGAGGCATTTGGCCATCATCTTCTTTGGGGACAACCCCAAGCTGAGTGA	584
Qy	580	GCTGTGTGGGGGCACATGCTCCACCCCTCTACATCCGGAGGTACATGTGGAGGCTGA	639
Db	585	GCTGTGTGGGGGCACCCGCTCACCCCTCTCATGTCCGGAGTGTACGTGTGAGCGCA	644
Qy	640	TTGGAGGCCCCCTGCCCTATCACGAGCCCGCTGTCTTTCCAAAGTACCTGACATAGA	699
Db	645	CGGCGGCCCCCTGCCCTATTACCGGCGCTCTGTGGAGACAGAGATATCAGAGATATGCA	704
Qy	700	CCAGATGATCTTCAACAGAGAGCGGCCCAAGGCTCTGGGCTCTCCACACATGTTGCTT	759
Db	705	GGAAGTATCTTCCACAGAGAGCTGCCCAAGCTTCTGAGACTCTCATACATGTTGCTT	764
Qy	760	CTACAGACACCGGGGAGGAACTGTGATCAATGACACAGGCTCCCGGTGTGCAATC	819
Db	765	CTACAAACACCAAGGAGACAAACCTGTCAAAATGACTACAGCCCCCGTGTGTGCAATC	824
Qy	820	AGGGAGCCGGGCCACCTGTGTTGGCTCTATCAACAATCTGGGGGCTGGTTCCT	879
Db	825	AGGGAGCCGGGCCACCTGTGTTGGCTCTATCAACAATCTGGGGGCTGGTTCCT	884
Qy	880	GCACACGTGGGGCTTGGAGCTGCTAGTGAACACACAAGGCCCTTGACCCCTGGAC	939
Db	885	TCACCCCTTGGCTTGGAGCTTCTGATAGATCAATAAGGCCCTGGAGTCTGGCTGTGAC	944
Qy	940	TATCCAGAAAGTGTCTATCAAGAGCCGCTCTACAGACAGCTGGGCCACTGGAGGCCA	999
Db	945	CATCCAGAAAGTATCTATCAAGAGCCGCTTACTATGAAAGTCTCATCTACCTGGAGGACCA	1004
Qy	1000	GTTTGGAGGCGGCTGTGATGTGTGTGATGCCAGAAATGCGCAGAGTGGGCTCG	1055
Db	1005	GTTTGGAGGCTGTGATGTGTGTGATGTGTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGT	1060
Qy	1060	GTTCCCTGAAGTCCCTGTGCCCGGGTCCAGCTCCCGCTCTACAGTTCTATCCCAAG	1119
Db	1065	GTTCTTAAAGTCCCACTGTCCACCGGGCCAGCTCCCGCTCTGTGACATGCCCAAGG	1124
Qy	1120	CCCCCGCTTGTGTGTCAGAGGAAGTGGAGGGCTCCTCACTGTGACATTTCTCTTTGG	1179
Db	1125	ACCCCGCTTGTGTGTCAGAGGAAGTGGAGGGCTCCTCACTGTGACATTTCTCTTTGG	1184
Qy	1180	CTGTGAGACATCTAGTGGCCCAAGATCTTTGACGTTCTGCTTCCAAAGAGGAAGTACGT	1239
Db	1185	CTGTGAGACATCTAGTGGCCCAAGATCTTTGATATCCGCTTTCAAAGGGAGAGGGGTGC	1244
Qy	1240	TTATGAGATAGCCTCCAAAGAGCCTTGGCCATCTATGTGTGAAAAATCCCAAGCAGCAAT	1299

Db	1245	CAATGAAATCAGATGTCACAGAGGCGCAATAGCCCTCTACGGGTGGAAAAATCCCGAGATCAAT	1304
QY	1300	GACGACCCGCGATGATGGATGGAGGCTTTGGCATAGGGCAAGTACACAGGCCCTGACCCG	1359
Db	1305	GTCCGACTGCTACGGGACGGGTACTTTGGCATTTGGCAAAATACTTACCCCTCGATCCG	1364
QY	1360	TGGGGTGGACTGCCCTACTTTGGCACCTACGTACGTGAGACTGGCACTTCCTTTTGGATGCCA	1419
Db	1365	AAGGGTAGACTGTCTTAACGTCCGCAACTATGTGAACTGGCACTTCCTTTTGGAAATCTCA	1424
QY	1420	GGCCCCCAAGCAATACGTGATGGCTTTTGTGTGTTGAACGAACAGGGCTCCCT	1479
Db	1425	GGCCCCCAAGCACTACGTGATGGCTTTTGTGTGTTGAACGAACAGGGCTCCCACT	1484
QY	1480	GGGGGACACACACAGATCGATACCTGCGCACTACTTTGGGGGTCTTGGGAAACGGTCT	1539
Db	1485	CGGGGCAACACACAGATCTTCACTCCCATTTTGGGGGTGTGTGGGACGGTCT	1544
QY	1540	GGTCGTGATCTATGTGCCACTTTGGCACTATGACTATGTGTGGATACGGTTTCCA	1599
Db	1545	TGTGCTGATTCGGTCTACTTTCCTGCTCAATTACGACTACATATGGGCAATGTCTTCCA	1604
QY	1600	CCCCAGTGGGCCATAGAAAATACGATTTCTATGCCACGGGCTACATCAGCTGGCAATCT	1659
Db	1605	CCCCAATGGGCCCATATACGCAAAATTCACGCCACAGGCTATACACTCAGCTTCTT	1664
QY	1660	CTTTGTGCTCTCTGTGGAAATACGGGAAACCAAGTGTACAGACACACCTTGGGACGGTCCA	1719
Db	1665	CTTCTGTCTCTGTAGAAATGTGGGAAACGAGTGTGGGGGCAACGCTGGGACGGTACA	1724
QY	1720	CACCCACAGCCCCCACTTCAAGGTGGATCGATGTAGACAGACTGGAGAACTGGGCTG	1779
Db	1725	CACCCACAGCGCTACTTCAAAATGGATCTGATGTGTGGAGGGCTGAAAGAACTGGGCTG	1784
QY	1780	GGCCGAGATATGCTTTTCTCCCATATGGCTGTGCCCTTGGAGCCCTGAGCACAGCTGCA	1839
Db	1785	GGCAGAGATATGGCTTTTCTCCCAAGATGTACTTGGCAACGGAGTACAGATGCA	1844
QY	1840	GAGGTGTAGGTGACCCGGAAAGCTGTGGAGATGAGAGAGACAGGCCCTCTCTGTGGG	1889
Db	1845	GAGGTGTAGGTGACTGGAAAGCTGTGGAGAGACAGGGAGGGGTGCTTCCCACTGGG	1904
QY	1900	AAGGCGCACCCCTGCTACCTGTACCTGGCGACGAACACACACAAGTGGGGTCAACC	1959
Db	1905	GGGGCGCACCCCACTACCTGTACCTGGCGAGTAAACACACACAAGTGGGGTCAAC	1964
QY	1960	CGGGGCTACCCGATCCAGATGCTCAGCTTTGTGTGAGAGCGGCTGCCCAACAGACTC	2019
Db	1965	GGGGGCTACCCGATCCAGATGCTTGTGTGAGAAAGCCCTTGCCCAAGAAATCC	2024
QY	2020	CATGGCGAGAGCTTACGCTGGGAGAGGTACACAGCTGGCTGTACCCAGCGGAAGAGGA	2079
Db	2025	CATAGAGAAAGCTTCACTCGGGGGGGAGTATCACTTGGCTGTGACCCAGAGAAAGAGGA	2084
QY	2080	GGAGCGCCAGTAGACGACGCTTTTCAATAGAAATACCCCTTGGGCGCCCACTGTGATTT	2139
Db	2085	GGAGCTACGACGCTTAGCATCTTCAACCAAGAACACCCGTGGAGCCCACTGTGGACTT	2144
QY	2140	CAGTACTTCATCAACATATAGACACTTGTCTGGAAGGATTTGGTGGCTGGGTGACAGC	2199
Db	2145	CACGACTTCATACGAATATAGACACTTGTCTGGAAGGACTTGGTATGCTGTGGTGTGACGC	2204
QY	2200	TGTTTTCTGCATATCCCAATGACAGAGGACATTTCTTAACAACAGTACTGTGGGGAACGG	2259
Db	2205	TGGCTTTTACACATCCCTTCATCACAANAATATCCCAACACGGTACTGGGGGAACCTC	2264
QY	2260	CGTGGGCTTCTCCGCAACCCATTAATCTTTTACGAABAACCCCTCTATACCTGTC	2319
Db	2265	AGTGGGCTTCTCCGCGCGGATTAATCTTTTACGAAGAACCCTCTTCCACTGTGTC	2324
QY	2320	CGATTCATCTACTTCCGAGGGGACAGAGATCTTGGGCTGTGCAAGTCAACCCCTAGC	2379
Db	2325	TGATTCATATTTCCGGAAGGGCCAGGATCTCACGGGCTGTAGAGTTAAACCCCTTGTG	238

Tue May 20 13:18:41 2003

us-10-081-408-1.rge

Page 27

[illegible]

Search completed: May 20, 2003, 06:42:46
Job time : 10509 secs

PR 25-SEP-2000: 2000US-234923P.
PR 25-SEP-2000: 2000US-234924P.
PR 25-SEP-2000: 2000US-235077P.
PR 25-SEP-2000: 2000US-235082P.
PR 25-SEP-2000: 2000US-235134P.
PR 25-SEP-2000: 2000US-235280P.
PR 26-SEP-2000: 2000US-235637P.
PR 26-SEP-2000: 2000US-235638P.
PR 27-SEP-2000: 2000US-235711P.
PR 27-SEP-2000: 2000US-235720P.
PR 27-SEP-2000: 2000US-235840P.
PR 27-SEP-2000: 2000US-235863P.
PR 28-SEP-2000: 2000US-236028P.
PR 28-SEP-2000: 2000US-236032P.
PR 28-SEP-2000: 2000US-236033P.
PR 28-SEP-2000: 2000US-236109P.
PR 28-SEP-2000: 2000US-236111P.
PR 29-SEP-2000: 2000US-236842P.
PR 29-SEP-2000: 2000US-236891P.
PR 02-OCT-2000: 2000US-237172P.
PR 02-OCT-2000: 2000US-237173P.
PR 02-OCT-2000: 2000US-237278P.
PR 02-OCT-2000: 2000US-237294P.
PR 02-OCT-2000: 2000US-237295P.
PR 02-OCT-2000: 2000US-237316P.
PR 03-OCT-2000: 2000US-237425P.
PR 03-OCT-2000: 2000US-237598P.
PR 03-OCT-2000: 2000US-237604P.
PR 03-OCT-2000: 2000US-237606P.
PR 01-NOV-2000: 2000US-244867P.
PR 01-NOV-2000: 2000US-245084P.
PA (AVALON PHARM.
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Sopfel DR, Weaver Z;
DR MPI; 2002-188264/24.
XX Screening for anti-neoplastic agent involves exposing cells to a
XX chemical agent to be tested for anti-neoplastic activity, and
XX determining a change in expression of a gene of a signature gene set -
XX
XX Claim 1; SEQ ID 241; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
XX anti-neoplastic agent. The method involves exposing cells to a chemical
XX agent to be tested for anti-neoplastic activity, determining a change in
XX expression of at least one gene (I) of a signature gene set, where (I)
XX comprises a sequence (S) selected from 8447 sequences (given in AB161664
XX to AB170110), or is at least 95% identical to (S), where a change in
XX expression is indicative of anti-neoplastic activity. (I) has cytostatic
XX activity and can be used in gene therapy. M1 can be used for screening
XX an anti-neoplastic agent, and can be used for producing a product which
XX is the data collected with respect to the anti-neoplastic agent as a
XX result of M1, and the data is sufficient to convey the chemical
XX structure and/or properties of the agent. M1 can be used in the
XX treatment of cancer such as colon, breast, stomach, lung, thyroid,
XX oesophageal, ovarian, kidney, prostate or pancreatic cancer,
XX adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
XX infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
XX carcinoma, papillary carcinoma and Wilms' tumor.
SQ Sequence 4040 BP; 825 A; 1183 C; 1097 G; 934 T; 1 other;
Query Match 100.0%; Score 4039.6; DB 24; Length 4040;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4039; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCCTCCACCTTACTGCCAGGATCTGACTACCGGGAACCTCAGCCAGAGTCCGGGA 60
|||||

Db 1 GTCCTCCACCTTACTGCCAGGATCTGACTACCGGGAACCTCAGCCAGAGTCCGGGA 60
QY 61 GCCCCACCCCGTCGACGAGCAACAGACCCCGTCTTCTGCGCTGAAATACATTG 120
|||||
Db 61 GCCCCACCCCGTCGACGAGCAACAGACCCCGTCTTCTGCGCGTGAATACATTG 120
QY 121 CTCTCTTTGGTGAATCAGCTGTCCCTCTTCTGCGGAAAATGAAACGAAACATCTT 180
|||||
Db 121 CTCTCTTTGGTGAATCAGCTGTCCCTCTTCTGCGGAAAATGAAACGAAACATCTT 180
QY 181 CGTCCCTCATCTCTGCGGCAATCAGCATCTTGGCTTGGTTGTTGTTGCTGCGGTTG 240
|||||
Db 181 CGTCCCTCATCTCTGCGGCAATCAGCATCTTGGCTTGGTTGTTGTTGCTGCGGTTG 240
QY 241 CAGGGGTGAGATGGGGGTAAACCCAGCCAGCTCCCATTTGCCCTCTATCTCCAG 300
|||||
Db 241 CAGGGGTGAGATGGGGGTAAACCCAGCCAGCTCCCATTTGCCCTCTATCTCCAG 300
QY 301 TGCCAGCTTTGACACACCCCTGGCCAGAGCCAGCTTTTTCAGACCTGAGCCGAGAGA 360
|||||
Db 301 TGCCAGCTTTGACACACCCCTGGCCAGAGCCAGCTTTTTCAGACCTGAGCCGAGAGA 360
QY 361 GCTGAGGCTGTGATGCGCTTTCGACCCAGCGGCTGGGGCAGAGGCTGTGATGCGAGC 420
|||||
Db 361 GCTGAGGCTGTGATGCGCTTTCGACCCAGCGGCTGGGGCAGAGGCTGTGATGCGAGC 420
QY 421 CCAGGCGCGGCGCTCGGACCAACTGTCTTCTCAGTGAAGTTGACGTGCTCCCAAGGC 480
|||||
Db 421 CCAGGCGCGGCGCTCGGACCAACTGTCTTCTCAGTGAAGTTGACGTGCTCCCAAGGC 480
QY 481 TGCAGCCCTGGCTCATTGGACAGGGGAGGAGCCCGCCAGCTGCGGAGGACATGGGCAT 540
|||||
Db 481 TGCAGCCCTGGCTCATTGGACAGGGGAGGAGCCCGCCAGCTGCGGAGGACATGGGCAT 540
QY 541 CGTCTTCTTTGGCAGGCAACCCAGCCCAAGTGAAGTGTGCTGCTGGGGCCACTGCC 600
|||||
Db 541 CGTCTTCTTTGGCAGGCAACCCAGCCCAAGTGAAGTGTGCTGCTGGGGCCACTGCC 600
QY 601 TCACCCCTCTATGCGGGAGCTGACTGTGAGCGCTCAGTGAAGAGCCCTGCTTATCA 660
|||||
Db 601 TCACCCCTCTATGCGGGAGCTGACTGTGAGCGCTCAGTGAAGAGCCCTGCTTATCA 660
QY 661 CCGAGCGCCGCTGCTGTTCCAGAGTACCTGACATGACCAATGATCTTCAACAGAGA 720
|||||
Db 661 CCGAGCGCCGCTGCTGTTCCAGAGTACCTGACATGACCAATGATCTTCAACAGAGA 720
QY 721 GCTGCCCGAGGCTTCTGAGCTTCTCCACACTGTGCTTCTACAGACCGGGAGCGGAA 780
|||||
Db 721 GCTGCCCGAGGCTTCTGAGCTTCTCCACACTGTGCTTCTACAGACCGGGAGCGGAA 780
QY 781 CTTGCTGACAAATGACACAGGCTCCCGTGTCTGCAATCAGAGGACCGGGCCACTGTGTT 840
|||||
Db 781 CTTGCTGACAAATGACACAGGCTCCCGTGTCTGCAATCAGAGGACCGGGCCACTGTGTT 840
QY 841 TGGGCTCTACACAAATCTGGGGCGGTGGTCTTCTGACAGCAGTGGGCTGGAGCT 900
|||||
Db 841 TGGGCTCTACACAAATCTGGGGCGGTGGTCTTCTGACAGCAGTGGGCTGGAGCT 900
QY 901 GCTAGTGAACCAAGGCGCTTGAACCTGCGCGGTGACATTCAGAGAGGTTCTATCA 960
|||||
Db 901 GCTAGTGAACCAAGGCGCTTGAACCTGCGCGGTGACATTCAGAGAGGTTCTATCA 960
QY 961 AGGCGGCTACTACGACAGCCTGGGCGGAGTGGAGCCAGTTTAGAGCGGCGCTGTGAA 1020
|||||
Db 961 AGGCGGCTACTACGACAGCCTGGGCGGAGTGGAGCCAGTTTAGAGCGGCGCTGTGAA 1020
QY 1021 TGTGCTGTGATCCAGCAATGCGACAGTGGGTGCTGCTGCTGAGTCCCTGTGCC 1080
|||||
Db 1021 TGTGCTGTGATCCAGCAATGCGACAGTGGGTGCTGCTGCTGAGTCCCTGTGCC 1080
QY 1081 CCGGGTCCAGCTCCCTTACAGTTCTATCCCAAGGCCCGGCTTCACTGTCCAGG 1140
|||||
Db 1081 CCGGGTCCAGCTCCCTTACAGTTCTATCCCAAGGCCCGGCTTCACTGTCCAGG 1140

Qy	1141	AAGTCAGTGGCCTCCTCACTGTGGACTTTCCTCTTGGCCTCGAGCAATTCAGTGGCC	1200
Db	1141	AAGTCAGTGGCCTCCTCACTGTGGACTTTCCTCTTGGCCTCGAGCAATTCAGTGGCC	1200
Qy	1201	AAGATTTTGAAGTTTGCCTCCAAAGAAAGACTAGTTTATGAGTAAACCTCCAGA	1260
Db	1201	AAGATCTTTGACGTTTGCCTCCAAAGAAAGACTAGTTTATGAGTAAACCTCCAGA	1260
Qy	1261	GGCCTTGGCATCTATGTGGAATTCGCCAGCAATGACGAGCCGCTATGTGATGG	1320
Db	1261	GGCCTTGGCATCTATGTGGAATTCGCCAGCAATGACGAGCCGCTATGTGATGG	1320
Qy	1321	AGGCTTTGGCATGGGCAAGTACACAGCCGCTGACCCGTGGGGGTGACTCCCTACTT	1380
Db	1321	AGGCTTTGGCATGGGCAAGTACACAGCCGCTGACCCGTGGGGGTGACTCCCTACTT	1380
Qy	1381	GGCCACCTACGTGGACTGGCACTTCCTTTTGGAGTCCAGAGGCCCCCAAGACAAATACGTGA	1440
Db	1381	GGCCACCTACGTGGACTGGCACTTCCTTTTGGAGTCCAGAGGCCCCCAAGACAAATACGTGA	1440
Qy	1441	TGCCCTTTGTGTGTTTGAACAGAACCAAGGGCCTCCCCCTGGGGGACACACTAGATCT	1500
Db	1441	TGCCCTTTGTGTGTTTGAACAGAACCAAGGGCCTCCCCCTGGGGGACACACTAGATCT	1500
Qy	1501	CTACCTCCACTACTTTTGGGGGCTTTCGGGAAAGCGTGGTGCATATGTGTCCAC	1560
Db	1501	CTACTCCCACTACTTTTGGGGGCTTTCGGGAAAGCGTGGTGCATATGTGTCCAC	1560
Qy	1561	CTTGCTCACTATGACTATGTGTGGATACGGTCTTCCACCCCACTGGGGCCATAGAAAT	1620
Db	1561	CTTGCTCACTATGACTATGTGTGGATACGGTCTTCCACCCCACTGGGGCCATAGAAAT	1620
Qy	1621	ACGATTTATGCCACGGGGTACATACAGTCGCGGCAATTCCTCTTTGGTGTACTGGGAAGTA	1680
Db	1621	ACGATTTATGCCACGGGGTACATACAGTCGCGGCAATTCCTCTTTGGTGTACTGGGAAGTA	1680
Qy	1681	CGGGAAACCAAGTGCACAGACACACCTGGGCAAGGTCACACCCACAGCCGCCCACTTCAA	1740
Db	1681	CGGGAAACCAAGTGCACAGACACACCTGGGCAAGGTCACACCCACAGCCGCCCACTTCAA	1740
Qy	1741	GGTGGATCTGGATGTATACAGAGACTGGAGAACTGGGCTGGGCGGAGATATAGCTTTGT	1800
Db	1741	GGTGGATCTGGATGTATACAGAGACTGGAGAACTGGGCTGGGCGGAGATATAGCTTTGT	1800
Qy	1801	CCCCATGGCTGTGCCCTGGAGCCCTTAGCAACCAAGCTGCAGAGGCTGCAGGTGACCCGGAA	1860
Db	1801	CCCCATGGCTGTGCCCTGGAGCCCTTAGCAACCAAGCTGCAGAGGCTGCAGGTGACCCGGAA	1860
Qy	1861	GCTGCTGGAGATGGAGAGCAGGCCCTTCTCTGTGGAGCGCCACCCCTCGCTACT	1920
Db	1861	GCTGCTGGAGATGGAGAGCAGGCCCTTCTCTGTGGAGCGCCACCCCTCGCTACT	1920
Qy	1921	GTACTGTGGCCAGCAACCAAGCAAGTGGGGTCAACCCCGGGGCTACCCGATCCAGAT	1980
Db	1921	GTACTGTGGCCAGCAACCAAGCAAGTGGGGTCAACCCCGGGGCTACCCGATCCAGAT	1980
Qy	1981	GCTCAGCTTGTCTGAGAGCCGCTGCCCAAAAAGACTTCATGGGAGAGGCTTCAGCTG	2040
Db	1981	GCTCAGCTTGTCTGAGAGCCGCTGCCCAAAAAGACTTCATGGGAGAGGCTTCAGCTG	2040
Qy	2041	GGAAAGGTACCACTGTGGCTGTACCCAAGGGAAGAGAGGCCCACTAGCAGCAGCT	2100
Db	2041	GGAAAGGTACCACTGTGGCTGTACCCAAGGGAAGAGAGGCCCACTAGCAGCAGCT	2100
Qy	2101	TTTCAATCAGATATACCTTTGGGCCCCCACTGTGATTCATCAACAATGA	2160
Db	2101	TTTCAATCAGATATACCTTTGGGCCCCCACTGTGATTCATCAACAATGA	2160
Qy	2161	GACCATTTGCTGGAAGGATTTGGTGGCTGGGTGACAGCTGTTTTCTGTCAATGCCACA	2220
Db	2161	GACCATTTGCTGGAAGGATTTGGTGGCTGGGTGACAGCTGTTTTCTGTCAATGCCACA	2220

QY	2221	TGCGAGAGCAATTCCTTAACACAGTACACTGTGGGGAAACGGCGTGGCTTCTCTCCGACC	2280
Db	2221	TGCGAGAGCAATTCCTTAACACAGTACACTGTGGGGAAACGGCGTGGCTTCTCTCCGACC	2280
QY	2281	CTATAACTCTTTAGAGAAACACCCCTCTTCTACCTGTGGCAGTCCATCTACTTCCGAGG	2340
Db	2281	CTATAACTCTTTAGAGAAACACCCCTCTTCTACTCTGCCACATCCATCTACTTCCGAGG	2340
QY	2341	GGACACGAGATCTGGGGGCTGCGAGGTCAACCCCTAGCTTGCCCTGCCCCAGGCTGTGC	2400
Db	2341	GGACACGAGATGTGGGGGCTGCGAGGTCAACCCCTAGCTTGCCCTGCCCCAGGCTGTGC	2400
QY	2401	CTGTGTCCCCCGGACCTCCCTGTCCCTTCTTCCCAAGGGGGCTTCTCTACACACTAGGGGGTCT	2460
Db	2401	CTGTGTCCCCCGGACCTCCCTGTCTCTTCCCAAGGGGGCTTCTCTCAACTAGGGGGTCT	2460
QY	2461	GGGATGGGGGATGTGGGCAAGGGGCTCCAGGGGCGAGGGTGTAGGGATGGGAGACGTGG	2520
Db	2461	GGGATGGGGGATGTGGGCAAGGGGCTCCAGGGGCGAGGGTGTAGGGATGGGAGACGTGG	2520
QY	2521	GCACGTGGGGCGGACGCGCTGGTTCCTCTTCTGTGTCGACGACCTCTTCTTCTCCACTAC	2580
Db	2521	GCACGTGGGGCGGAGCGCTGGTTCCTCTTCTGTGTCGACGACCTCTTCTTCTCCACTAC	2580
QY	2581	CCTCCCTCCGATCCGCGCTCTGAGCCAGAGACCTCTGACACCTCTATCTCCCTGACACAGG	2640
Db	2581	CCTCCCTCCGATCCGCGCTCTGAGCCAGAGACCTCTGACACCTCTGATCCCTGACACAGG	2640
QY	2641	GACACTGAACTGTGTGATGTCGAGCGTACTAGTTCATCCACAGAGGGCAGGACATGG	2700
Db	2641	GACACTGAACTGTGTGATGTCGAGCGTACTAGTTCATCCACAGAGGGCAGGACATGG	2700
QY	2701	CCGACGCTGGAGCCGCTGGGCGAGGGGCTCCAGATGTTCCCTTGTGTGCTGTGCTGGCT	2760
Db	2701	CCGACGCTGGAGCCGCTGGGCGAGGGGCTCCAGATGTTCCCTTGTGTGCTGTGCTGGCT	2760
QY	2761	TTCCCGAATCTTTTAGGCGCACTCCAGAGACTCTAAAGGGGGCTATTCCCTGGAGAC	2820
Db	2761	TTCCCGAATCTTTTAGGCGCACTCCAGAGACTCTAAAGGGGGCTATTCCCTGGAGAC	2820
QY	2821	CCAGAGTAGGGTTGTCGAGTCCCTGCAAGTGCATAGCTGACGTGGAAAGGATGCTGTGCTC	2880
Db	2821	CCAGAGTAGGGTTGTCGAGTCCCTGCAAGTGCATAGCTGACGTGGAAAGGATGCTGTGCTC	2880
QY	2881	ACATTCCCTCTCATCCAGAGTCCCTTCTCTGCTTCCCTCTCTCTACACTACTTCCCTCC	2940
Db	2881	ACATTCCCTCTCATCCAGAGTCCCTTCTCTGCTTCCCTCTCTCTCTACACTACTTCCCTCC	2940
QY	2941	TCTCTCTCTGTTCTGCTCCCTTCTCTTCTATCTGCAATTTCTCCGAAATCTGTAGGGGAT	3000
Db	2941	TCTCTCTCTCTGTTCTGCTCCCTTCTCTTCTATCTGCAATTTCTCCGAAATCTGTAGGGGAT	3000
QY	3001	ATCCCTATGTCTCCAGCCCCCTGGTACTGCCCAACCCCTCAGTTTCAAGTCAAGTTCCTCT	3060
Db	3001	ATCCCTATGTCTCCAGCCCCCTGGTACTGCCCAACCCCTCAGTTTCAAGTCAAGTTCCTCT	3060
QY	3061	CCTCTCCAGCCCTATGGAAGTCTCAAGAGTACAGGGGACCCCTATAGAGTGGCCAAATCC	3120
Db	3061	CCTCTCTCCAGCCCTATGGAAGTCTCAAGAGTACAGGGGACCCCTAATAGAGTGGCCAAATCC	3120
QY	3121	TGTGTGTGTGTCCTTGTGTCTGTGTCTTATTGGGAGTAGGAGTTGTCTCTACCCCTGTGC	3180
Db	3121	TGTGTGTGTGTCCTTGTGTCTGTGTCTTATTGGGAGTAGGAGTTGTCTCTACCCCTGTGC	3180
QY	3181	CTGGGGGCTGGGTGTTCACAGACGTGCTTGTGCAATTTGTGTCTGCTGCTCATG	3240
Db	3181	CTGGGGGCTGGGTGTTCACAGACGTGCTTGTGCAATTTGTGTCTGCTGCTCATG	3240
QY	3241	CTCTCTATAGAGGAGATGTGATCGAGACAGACAGCTCAAGTTAGCATTTCAAGTGA	3300
Db	3241	CTCTCTATAGAGGAGATGTGATCGTGCACAGCAGACGCTCAAGTTAGCATTTCAAGTGA	3300
QY	3301	TTTGGGGGTGCATGATGAAGAAATGGCCATTTTGTACCAAGGGGCTGTGATTTCTGC	3360


```
Db 3301 TTGGGGGTGCAATGATATATGAGAAATGGCCATTTTGTACAGGGCTCTGTATTTCTGCA 3360
|||
Qy 3361 CAGCCTTTGGAGAGCTGAGTGAACAAAGGGTGGGATCAAGATGAGAACCCAA 3420
|||
Db 3361 CAGCCTTTGGAGAGCTGAGTGAACAAAGGGTGGGATCAAGATGAGAACCCAA 3420
|||
Qy 3421 GCGCCTTCAAACTCCAGCCACCACAGAGAGGGCTGCCAATACATTCACGATGCGA 3480
|||
Db 3421 GCGCCTTCAAACTCCAGCCACCACAGAGAGGGCTGCCAATACATTCACGATGCGA 3480
|||
Qy 3481 ATGAGCTGGGGCCCTGGGTGAGGTGGGGGTCTGGCTAGTGGGGAGGGGCTGGGCTGGGT 3540
|||
Db 3481 ATGAGCTGGGGCCCTGGGTGAGGTGGGGGTCTGGCTAGTGGGGAGGGGCTGGGCTGGGT 3540
|||
Qy 3541 GGGGAGAGGGCTGGGCTGGTCCAGAGCTTGGGCTCCATTCACATCTGCTCCCTCTG 3600
|||
Db 3541 GGGGAGAGGGCTGGGCTGGTCCAGAGCTTGGGCTCCATTCACATCTGCTCCCTCTG 3600
|||
Qy 3601 AGGCTGTGATTTGGGATGGGAGCAAAAGAAATAGCAAGAGATGAGAAACAGAACTTT 3660
|||
Db 3601 AGGCTGTGATTTGGGATGGGAGCAAAAGAAATAGCAAGAGATGAGAAACAGAACTTT 3660
|||
Qy 3661 TTTCCTAAAGAGCTGGTAAATCAATCTGATACAGCCTTACAAATACATAGTATGCGAG 3720
|||
Db 3661 TTTCCTAAAGAGAGCTGGTAAATCAATCTGATACAGCCTTACAAATAGTATGCGAG 3720
|||
Qy 3721 CTAATAAATAATTTGATGCTTTATATATATATATATATATATATATATATATATATAT 3780
|||
Db 3721 CTAATAAATAATTTGATGCTTTATATATATATATATATATATATATATATATATATAT 3780
|||
Qy 3781 AAGCCACAGAAATGATGTATACGACCTCCATCTTGTCTTACAGAAAGAGATGAAATTA 3840
|||
Db 3781 AAGCCACAGAAATGATGTATACGACCTCCATCTTGTCTTACAGAAAGAGATGAAATTA 3840
|||
Qy 3841 AACACATATATGCTATATATATATATATATATATATATATATATATATATATATATAT 3900
|||
Db 3841 AACACATATATGCTATATATATATATATATATATATATATATATATATATATATATAT 3900
|||
Qy 3901 GAGGGGAAACGATAGTTGAGAGACAGAGAGGAAAGGCTTTACACTTACACCTTTTG 3960
|||
Db 3901 GAGGGGAAACGATAGTTGAGAGACAGAGAGGAAAGGCTTTACACTTACACCTTTTG 3960
|||
Qy 3961 TACATTTTGAATTTTGAACATGATGATGATATACATTAATTAATTAATTAATGAG 4020
|||
Db 3961 TACATTTTGAATTTTGAACATGATGATGATATACATTAATTAATTAATTAATGAG 4020
|||
Qy 4021 CCAAAAAAAAAAAAAAAAAA 4040
|||
Db 4021 CCAAAAAAAAAAAAAAAAAA 4040
|||
RESULT 2
ABL65211
ID ABL65211 standard; DNA: 4040 BP.
XX
AC ABL65211:
XX
DT 15-MAY-2002 (first entry)
XX
DE Lung cancer related gene sequence SEQ ID NO:3548.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
```

```
PF 30-MAY-2001; 2001WO-US10838.
XX
XX 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-231331P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235027P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237177P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237315P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
XX (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
XX WPI: 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
XX chemical agent to be tested for anti-neoplastic activity, and
XX determining a change in expression of a gene of a signature gene set
XX
XX Claim 1; SEQ ID 3548; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
XX anti-neoplastic agent. The method involves exposing cells to a chemical
XX agent to be tested for anti-neoplastic activity, determining a change in
XX expression of at least one gene (I) of a signature gene set, where (I)
XX comprises a sequence (S) selected from 8447 sequences (given in ABL61664
XX to ABL70110), or is at least 95% identical to (S), where a change in
XX expression is indicative of anti-neoplastic activity. (I) has cytostatic
XX activity and can be used in gene therapy. M1 can be used for screening
XX an anti-neoplastic agent, and can be used for producing a product which
XX is the data collected with respect to the anti-neoplastic agent as a
XX result of M1, and the data is sufficient to convey the chemical
XX structure and/or properties of the agent. M1 can be used in the
XX treatment of cancer such as colon, breast, stomach, lung, thyroid,
XX oesophageal, ovarian, kidney, prostate or pancreatic cancer,
XX adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
```

CC Infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilm's tumour.

XX Sequence 4040 BP; 825 A; 1183 C; 1097 G; 934 T; 1 other:

Query Match 100.0%; Score 4039.6; DB 24; Length 4040;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4039; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCTTCCACCCCTAGTCCAGGACATGACTACCGGGAACCTCAGCCAGAGTCCGGA 60
DB 1 GTCTTCCACCCCTAGTCCAGGACATGACTACCGGGAACCTCAGCCAGAGTCCGGA 60
QY 61 GCGCCCGACCCCGTCCAGAGGACAGAGGCGCCCTTGTGCGGTGAGATTAATG 120
DB 61 GCGCCCGACCCCGTCCAGAGGACAGAGGCGCCCGTGTGCGGTGAGATTAATG 120
QY 121 CTCTCTTTGGTGAATCAAGCTGTCCCTCTTGTGGAAATGAACAGAAATCAATCT 180
DB 121 CTCTCTTTGGTGAATCAAGCTGTCCCTCTTGTGGAAATGAACAGAAATCAATCT 180
QY 181 CGTGTCTCATCTGCGCGTCATCAACATCTTGGCTTGGTTGTCTCTGCTGTGG 240
DB 181 CGTGTCTCATCTGCGCGTCATCAACATCTTGGCTTGGTTGTCTCTGCTGTGG 240
QY 241 CAGGGGTGAGATGGGGGTGAACCCAGCCAGCTTCCCATGCGCCCTGTATCTCCAG 300
DB 241 CAGGGGTGAGATGGGGGTGAACCCAGCCAGCTTCCCATGCGCCCTGTATCTCCAG 300
QY 301 TGCCAGACCTTGGACACACCTGTGCGAGAGCAGCTTTGGACAGCTTGAGCCGAGAGA 360
DB 301 TGCCAGACCTTGGACACACCTGTGCGAGAGCAGCTTTGGACAGCTTGAGCCGAGAGA 360
QY 361 GCTGAGGCTGTATGCGCTTTCTGACCCAGCGGCTGGGGCCAGGCGCTGTGATGAGC 420
DB 361 GCTGAGGCTGTATGCGCTTTCTGACCCAGCGGCTGGGGCCAGGCGCTGTGATGAGC 420
QY 421 CGAGGCGCGGCGCTCGGACAACTGTGCTCTGAGTGGATTCAGAGTGGCTCCCAAGC 480
DB 421 CGAGGCGCGGCGCTCGGACAACTGTGCTCTGAGTGGATTCAGAGTGGCTCCCAAGC 480
QY 481 TGCAGCCCTGTGCTACTTGGACAGGGGAGGAGCCGCCACTGCGCGGAGGACATG 540
DB 481 TGCAGCCCTGTGCTACTTGGACAGGGGAGGAGCCGCCACTGCGCGGAGGACATG 540
QY 541 CGTCTTCTTTGGAGGCAACCCAGGCCAAGTGAAGTGAAGTGTGGGGCCACTGCC 600
DB 541 CGTCTTCTTTGGAGGCAACCCAGGCCAAGTGAAGTGAAGTGTGGGGCCACTGCC 600
QY 601 TCACCCCTCTATATGCGGGAGCTGAGTGGAGCGTCAATGAGAGGCGCCCTGCTATCA 660
DB 601 TCACCCCTCTATATGCGGGAGCTGAGTGGAGCGTCAATGAGAGGCGCCCTGCTATCA 660
QY 661 CCGAGCGCCCGCTGTCTCCAGAGTACCTGAGATAGACAGATGAATCTTCAACAGAGA 720
DB 661 CCGAGCGCCCGCTGTCTCCAGAGTACCTGAGATAGACAGATGAATCTTCAACAGAGA 720
QY 721 GCTGCGCCAGGCTTGTGGGCTTCTCACCACCTGTGCTTACAAAGCACGGGAGC 780
DB 721 GCTGCGCCAGGCTTGTGGGCTTCTCACCACCTGTGCTTACAAAGCACGGGAGC 780
QY 781 CCGTGTGAAATGACACAGGCTCCCGTGTGCAATGAGGGAACCGGCGCACTGCTT 840
DB 781 CCGTGTGAAATGACACAGGCTCCCGTGTGCAATGAGGGAACCGGCGCACTGCTT 840
QY 841 TGCGCTCTACTACATCTCGGGCTGTGCTTCTCTGACACAGCTGTGGGCTGTGAGCT 900
DB 841 TGCGCTCTACTACATCTCGGGCTGTGCTTCTCTGACACAGCTGTGGGCTGTGAGCT 900
QY 901 GCTAGTGAACACAGGCGCTTACCGTGCCTGAGTATCCAGAAAGTGTCTATYCA 960
DB 901 GCTAGTGAACACAGGCGCTTACCGTGCCTGAGTATCCAGAAAGTGTCTATYCA 960

QY 961 AGGCGGCTACTAGACAGCCGTGGCCAGCGTGAAGGCGCCAGTTTGAGGGCCGCGCTGTGAA 1020
DB 961 AGGCGGCTACTAGACAGCCGTGGCCAGCGTGAAGGCGCCAGTTTGAGGGCCGCGCTGTGAA 1020
QY 1021 TGTGTGCTGATCCACAGCAATGGACAGGTGGGTCTGTGCTTCCCTGAAGTCCCTGTGCC 1080
DB 1021 TGTGTGCTGATCCACAGCAATGGACAGGTGGGTCTGTGCTTCCCTGAAGTCCCTGTGCC 1080
QY 1081 CCGGCGTCAAGCTCCCTCTTACATCTTATCCCAAGGCGCCCGCTTCAAGTCTCAGGG 1140
DB 1081 CCGGCGTCAAGCTCCCTCTTACATCTTATCCCAAGGCGCCCGCTTCAAGTCTCAGGG 1140
QY 1141 AAGTCGATGGCTCCCTCACTGTGAGTCTTCTGCTGCGTGGAGACTTCAATGAGCC 1200
DB 1141 AAGTCGATGGCTCCCTCACTGTGAGTCTTCTGCTGCGTGGAGACTTCAATGAGCC 1200
QY 1201 AAGATCTTTGAGCTTCCAGAGAGAAAGCTAGTTATGATTAAGCTTCAAGA 1260
DB 1201 AAGATCTTTGAGCTTCCAGAGAGAAAGCTAGTTATGATTAAGCTTCAAGA 1260
QY 1261 GCGCTTGGCATATATGTTGAATTCGCCAGCAGTAAGCAGCCGCTATGTGATGG 1320
DB 1261 GCGCTTGGCATATATGTTGAATTCGCCAGCAGTAAGCAGCCGCTATGTGATGG 1320
QY 1321 AGGCTTGGCATGGGCAAGTACACAGGCGCCCTGACCCGTTGGATGACTGCCCTACTT 1380
DB 1321 AGGCTTGGCATGGGCAAGTACACAGGCGCCCTGACCCGTTGGATGACTGCCCTACTT 1380
QY 1381 GCGCACCTAGCTGAGTGGAGCTTCTTTGAGTCCAGGCGCCCAAGCAATATACATGA 1440
DB 1381 GCGCACCTAGCTGAGTGGAGCTTCTTTGAGTCCAGGCGCCCAAGCAATATACATGA 1440
QY 1441 TGCCCTTTGTGTGTTGAACAGAAACAGGCGCTCCCTGTGGCGGACACATCAGATCT 1500
DB 1441 TGCCCTTTGTGTGTTGAACAGAAACAGGCGCTCCCTGTGGCGGACACATCAGATCT 1500
QY 1501 CTACTGCACTACTTGTGGGGTCTTGGGAAAGGCTGTGCTGTGATATGATGAC 1560
DB 1501 CTACTGCACTACTTGTGGGGTCTTGGGAAAGGCTGTGCTGTGATATGATGAC 1560
QY 1561 CTGCTCACTATGATGTGTGGATAGGGTCTTCCACCCAGTGGGGCATAGAAAT 1620
DB 1561 CTGCTCACTATGATGTGTGGATAGGGTCTTCCACCCAGTGGGGCATAGAAAT 1620
QY 1621 ACGATTTATGCCAGCGGCTATCATCAGCTGGGCAATCTCTTGTGTGATGGAAGTA 1680
DB 1621 ACGATTTATGCCAGCGGCTATCATCAGCTGGGCAATCTCTTGTGTGATGGAAGTA 1680
QY 1681 CCGGAAACAAAGTGCAGAGCAGCACCTGTGGCAGGTCACACCCAGCGGCCACTTCAA 1740
DB 1681 CCGGAAACAAAGTGCAGAGCAGCACCTGTGGCAGGTCACACCCAGCGGCCACTTCAA 1740
QY 1741 GGTGATCTGATGTAGCAGAGACTGGAACCTGGGCTGTGGCCGAGATATGCTTTGT 1800
DB 1741 GGTGATCTGATGTAGCAGAGACTGGAACCTGGGCTGTGGCCGAGATATGCTTTGT 1800
QY 1801 CCCGATGGCTGTGCTGTGAGCCCTGAGCACACCTGCAAGAGCTGCAAGTGCACCGGAA 1860
DB 1801 CCCGATGGCTGTGCTGTGAGCCCTGAGCACACCTGCAAGAGCTGCAAGTGCACCGGAA 1860
QY 1861 GGTGCTGGAATGAGAGGAGGAGGCGGCTTCTGCTGGGAAAGGCGCCCTGCTACT 1920
DB 1861 GGTGCTGGAATGAGAGGAGGAGGCGGCTTCTGCTGGGAAAGGCGCCCTGCTACT 1920
QY 1921 GTACTGTGGCAGCAACACAGCAACAAAGTGGGCTACACCCCGGGGCTTACCGCATCAGAT 1980
DB 1921 GTACTGTGGCAGCAACACAGCAACAAAGTGGGCTACACCCCGGGGCTTACCGCATCAGAT 1980
QY 1981 GCTCAGCTTTGCTGTGAGAGCGCGTGGCCCAAAACAGCTCATGTGCGAGAGCTTCACTG 2040
DB 1981 GCTCAGCTTTGCTGTGAGAGCGCGTGGCCCAAAACAGCTCATGTGCGAGAGCTTCACTG 2040
QY 2041 GGAAGGTACACAGCTGTGAGACCCAGCGGAAGAGAGGAGGCGGATGACACAGCGT 2100

Db 2041 GGAGAGGTACCAGCTGGCTGTGACCCAGCGGAAGAGAGAGAGCCCGCTAGCAGACGCT 2100
Qy 2101 TTTCATCAGAAATGACCCCTTGGGCCCCACTGTGATTTCACTGACTTCAACAATGA 2160
Db 2101 TTTCATCAGAAATGACCCCTTGGGCCCCACTGTGATTTCACTGACTTCAACAATGA 2160
Qy 2161 GACCATTTGCTGGAAGATTTGGTGGCTGGGTGACAGCTGGTGTTCATATCCACA 2220
Db 2161 GACCATTTGCTGGAAGATTTGGTGGCTGGGTGACAGCTGGTGTTCATATCCACA 2220
Qy 2221 TGCAGAGACATTCCTTAACACAGTACTGGGGAAGGGGTGGCTTCTCCCGAGC 2280
Db 2221 TGCAGAGACATTCCTTAACACAGTACTGGGGAAGGGGTGGCTTCTCCCGAGC 2280
Qy 2281 CTATTAATCTTTGACGAGAACCCCTTCTACTGTCCGACTCATCTATCCAGG 2340
Db 2281 CTATTAATCTTTGACGAGAACCCCTTCTACTGTCCGACTCATCTATCCAGG 2340
Qy 2341 GGACCGAGATGTGGGGGCGGAGGTCACACCCCTAGCTGGCTGGCCCGAGGCTGCTG 2400
Db 2341 GGACCGAGATGTGGGGGCGGAGGTCACACCCCTAGCTGGCTGGCCCGAGGCTGCTG 2400
Qy 2401 CTGTGCCCCGACCTCCCTGCTCTCTCCACGCGGGCTTCTCTACAACTAGCGGTCT 2460
Db 2401 CTGTGCCCCGACCTCCCTGCTCTCTCCACGCGGGCTTCTCTACAACTAGCGGTCT 2460
Qy 2461 GGGATGGGGCATGTGGCCAAAGGCTCCAGGGCCAGGGTGTGAGGGATGGGGAGCAGCTGG 2520
Db 2461 GGGATGGGGCATGTGGGGCAAGGGCTCCAGGGCCAGGGTGTGAGGGATGGGGAGCAGCTGG 2520
Qy 2521 GCACCTGGGGCGGACGCTGTGCTTCCCTTTCCTGTGCGCAGAGACTCTTTTCTTCCACTAC 2580
Db 2521 GCACCTGGGGCGGACGCTGTGCTTCCCTTTCCTGTGCGCAGAGACTCTTTTCTTCCACTAC 2580
Qy 2581 CCTCCCTCGGATCGGCTCTGAGCCAGAGAGGCTCTGACCTGTGATGCTGACACAGAG 2640
Db 2581 CCTCCCTCGGATCGGCTCTGAGCCAGAGAGGCTCTGACCTGTGATGCTGACACAGAG 2640
Qy 2641 GACACTGAACCTTGTGATGCGCAGCTGACTGAGTTCTCATCAGAGAGGCGAGGATGG 2700
Db 2641 GACACTGAACCTTGTGATGCGCAGCTGACTGAGTTCTCATCAGAGAGGCGAGGATGG 2700
Qy 2701 CCAGGCTGAGACCGCTGGGAGGGCTTCCCTAATGATGGTTCCCTTGTGTGCTGCT 2760
Db 2701 CCAGGCTGAGACCGCTGGGAGGGCTTCCCTAATGATGGTTCCCTTGTGTGCTGCT 2760
Qy 2761 TTCCGATCTTTTATAGGCACTCTCAAGGACTATAAAGGGGCTATTCCCTGGAGAC 2820
Db 2761 TTCCGATCTTTTATAGGCACTCTCAAGGACTATAAAGGGGCTATTCCCTGGAGAC 2820
Qy 2821 CCAGAGTAGGGTTGCCAGTCTGCAAGTCCATAGCTGAGTGAAGAGATCTTCTGCTC 2880
Db 2821 CCAGAGTAGGGTTGCCAGTCTGCAAGTCCATAGCTGAGTGAAGAGATCTTCTGCTC 2880
Qy 2881 ACATTCCTCTCATCGAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2940
Db 2881 ACATTCCTCTCATCGAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2940
Qy 2941 TCT 3000
Db 2941 TCT 3000
Qy 3001 ATCCCTATGTCAGAGCCCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3060
Db 3001 ATCCCTATGTCAGAGCCCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3060
Qy 3061 CCTCTCAGCCCTATGGAAGCTCAAGGCTCAGGGAGCCCTTAATCAGAGTGGCCATCC 3120
Db 3061 CCTCTCAGCCCTATGGAAGCTCAAGGCTCAGGGAGCCCTTAATCAGAGTGGCCATCC 3120
Qy 3121 TGTGTGTCT 3180
Db 3121 TGTGTGTCT 3180

Db 3121 TGTGTGTCT 3180
Qy 3181 CTGGGGCTGGGGTGTGTTTCAGAGACGCTGCTCTGTGACATTTGTGTGCTGCTGCTATG 3240
Db 3181 CTGGGGCTGGGGTGTGTTTCAGAGACGCTGCTCTGTGACATTTGTGTGCTGCTGCTATG 3240
Qy 3241 CTCTCTATAGAGAGATGGTGTGATCTGTCAGACAGCAGCTCAAGTATGCAATTCAGTGA 3300
Db 3241 CTCTCTATAGAGAGATGGTGTGATCTGTCAGACAGCAGCTCAAGTATGCAATTCAGTGA 3300
Qy 3301 TTTGGGGGTGCAATGATTAATGAAGAAATGCGCAATTTTGTACAGAGGCTGTGATTTCTGCA 3360
Db 3301 TTTGGGGGTGCAATGATTAATGAAGAAATGCGCAATTTTGTACAGAGGCTGTGATTTCTGCA 3360
Qy 3361 CAGCTCTTTGGAGGCTGAGTGAAGCAAAAGGCTGGCATCAAGATGAGAACCAAA 3420
Db 3361 CAGCTCTTTGGAGGCTGAGTGAAGCAAAAGGCTGGCATCAAGATGAGAACCAAA 3420
Qy 3421 GCGCCCTACAACTCCAGCACCCAGCCAGAGAGGGGCTGTCCAAATCAATAGGATGCGA 3480
Db 3421 GCGCCCTACAACTCCAGCACCCAGCCAGAGAGGGGCTGTCCAAATCAATAGGATGCGA 3480
Qy 3481 ATGAGCTGGGGCTGGGTGAGGTGGGGTCTGGGCTAGTGGGAGGGGCTGGGCTGGGT 3540
Db 3481 ATGAGCTGGGGCTGGGTGAGGTGGGGTCTGGGCTAGTGGGAGGGGCTGGGCTGGGT 3540
Qy 3541 GGGGAGGGGCTGGGCTGGGTGAGGTGGGGTCTGGGCTAGTGGGAGGGGCTGGGCTGGGT 3600
Db 3541 GGGGAGGGGCTGGGCTGGGTGAGGTGGGGTCTGGGCTAGTGGGAGGGGCTGGGCTGGGT 3600
Qy 3601 AGGCTGATGGGGGATGGGGGACAAAGAAATAGCAAGATGAGAAACAGAACTTT 3660
Db 3601 AGGCTGATGGGGGATGGGGGACAAAGAAATAGCAAGATGAGAAACAGAACTTT 3660
Qy 3661 TTTCTTAAGGACCTGTTAATCAATTCATGATCAGCTTACATCAATAGATATGACAG 3720
Db 3661 TTTCTTAAGGACCTGTTAATCAATTCATGATCAGCTTACATCAATAGATATGACAG 3720
Qy 3721 CTAAAAAATTAATGATGCTCTTATATATATATATATATATATATATATATATATATAT 3780
Db 3721 CTAAAAAATTAATGATGCTCTTATATATATATATATATATATATATATATATATATAT 3780
Qy 3781 AAGGACAGAAATGATATAGGACGACCTCCATTTGTTTCAAGAGGATATATATA 3840
Db 3781 AAGGACAGAAATGATATAGGACGACCTCCATTTGTTTCAAGAGGATATATATA 3840
Qy 3841 AACACATTAATGCTATGATATGCTATATTCAGAAATTAATGGGTAACAGTATTTTGG 3900
Db 3841 AACACATTAATGCTATGATATGCTATATTCAGAAATTAATGGGTAACAGTATTTTGG 3900
Qy 3901 GAGGGGAACAGTAGGTTGAGAGACAGAGAGAGGAGGCTTTAACTACATTCACCTTTTG 3960
Db 3901 GAGGGGAACAGTAGGTTGAGAGACAGAGAGAGGAGGCTTTAACTACATTCACCTTTTG 3960
Qy 3961 TACATTTGAATTTTGAACCTGACGATTAATCTATTCATTAATTAATTAATGGG 4020
Db 3961 TACATTTGAATTTTGAACCTGACGATTAATCTATTCATTAATTAATTAATGGG 4020
Qy 4021 CCAAAAAAAAAAAAAAAAAA 4040
Db 4021 CCAAAAAAAAAAAAAAAAAA 4040

RESULT 3
AAAX28632
ID AAX28632 standard; DNA; 2501 BP.
XX AAX28632;
AC AAX28632;
XX 21-JUN-1999 (first entry)
XX Nucleotide sequence of the vascular adhesion protein-1.
DE XX

Oy	621	ACGGAATGTGGAGCGCATAGAGAGGCCCTCCCTATACCGAGAGCCCTGCTGTTC	680
Db	541	ACGGAATGTGGAGCGCATAGAGAGGCCCTCCCTATACCGAGAGCCCTGCTGTTC	600
Oy	681	AAGAGTACCTGGACATAGACACAGATATCTTAACAGAGAGCTGGCCACGCTCTGAGC	740
Db	601	AAGAGTACCTGGACATAGAGACAGATATCTTAACAGAGAGCTGGCCACGCTCTGAGC	660
Oy	741	TTCTCCACCACTGTTCCTTCTACAMGACACCGGGGACGAACTGGTGAAATGACACGG	800
Db	661	TTCTCCACCACTGTTCCTTCTACAMGACACCGGGGACGAACTGGTGAAATGACACGG	720
Oy	801	CTCCCGCTGGTCTGCATATAGAGGAGACCGGGGACACCGGGTTGGGGCTCTACTACAAACACT	860
Db	721	CTCCCGCTGGTCTGCATATAGAGGAGACCGGGGACACCGGGTTGGGGCTCTACTACAAACACT	780
Oy	861	CGGGCGCTGGGTTCTTCTTCGACACACGTCGGGCTTGGAGCTGCTACTGAACACAGAGGCC	920
Db	781	CGGGCGCTGGGTTCTTCTTCGACACACGTCGGGCTTGGAGCTGCTACTGAACACAGAGGCC	840
Oy	921	TTGACCCCTGGCCGCTGGACATACAGAAAGTCTTCTATCAAGCCGCTACTACAGACACC	980
Db	841	TTGACCCCTGGCCGCTGGACATACAGAAAGTCTTCTATCAAGCCGCTACTACAGACACC	900
Oy	981	TGGCCCACTGGAGGCGCCAGTTTGAAGCCGGCGCTGGTGATTTGGTCTGATCCACAGA	1040
Db	901	TGGCCCACTGGAGGCGCCAGTTTGAAGCCGGCGCTGGTGATTTGGTCTGATCCACAGA	960
Oy	1041	ATGGCACAGGTGGGTCTGTGCTCCCTGTAAGTCCCTCTGTGCCCCGGGCTCAGCTCCCCCTC	1100
Db	961	ATGGCACAGGTGGGTCTGTGCTCCCTGTAAGTCCCTCTGTGCCCCGGGCTCAGCTCCCCCTC	1020
Oy	1101	TACAGTTCTATCCCAAGAGGCCCGCGCTTGAAGTCCAGGGAAGTGAAGGCGTCCTAC	1160
Db	1021	TACAGTTCTATCCCAAGAGGCCCGCGCTTGAAGTCCAGGGAAGTGAAGGCGTCCTAC	1080
Oy	1161	TGTGGACTTTCCTCTTTGGGCTCTGGAGCAATTCAGTGGCCCAAGATCTTTGACGTTGCT	1220
Db	1081	TGTGGACTTTCCTCTTTGGGCTCTGGAGCAATTCAGTGGCCCAAGATCTTTGACGTTGCT	1140
Oy	1221	TCCAAAGAGAAAGACTAGTTTATGAATTAAGGCTCCAAAGGCCCTGGCCATCTATGGTG	1280
Db	1141	TCCAAAGAGAAAGACTAGTTTATGAATTAAGGCTCCAAAGGCCCTGGCCATCTATGGTG	1200
Oy	1281	GAATTTCCCGACGACAAATAGACACCCCGCTATGTGATGGAGGCTTTGGCATGGGCAAGT	1340
Db	1201	GAATTTCCCGACGACAAATAGACACCCCGCTATGTGATGGAGGCTTTGGCATGGGCAAGT	1260
Oy	1341	ACACACAGCCCTCGAACCCCGGGGGTGGAGATGGCCCTACTTGGCCACCTACGTGGACTGGC	1400
Db	1261	ACACACAGCCCTCGAACCCCGGGGGTGGAGATGGCCCTACTTGGCCACCTACGTGGACTGGC	1320
Oy	1401	ACTTCTCTTTTGGAGTCCACAGGCCCCCAACAAATAAGTATCCCTTTGTGTTGTGAAC	1460
Db	1321	ACTTCTCTTTTGGAGTCCACAGGCCCCCAACAAATAAGTATCCCTTTGTGTTGTGAAC	1380
Oy	1461	AGAACACAGGGCTCCCTCTCGGGCGACACCACTCAGATCTTACTGCGACTTCTTTGGG	1520
Db	1381	AGAACACAGGGCTCCCTCTCGGGCGACACCACTCAGATCTTACTGCGACTTCTTTGGG	1440
Oy	1521	GTCCTGGGGAAGAGGTCGTGGTGTGAGATATATGTCACCTTGTCAACTATGACTATG	1580
Db	1441	GTCCTGGGGAAGAGGTCGTGGTGTGAGATATATGTCACCTTGTCAACTATGACTATG	1500
Oy	1581	TGTGGGATACGGTCTTCCACCCCACTGGGGCATAAGAAATACGATTCATGCAAGGGCT	1640
Db	1501	TGTGGGATACGGTCTTCCACCCCACTGGGGCATAAGAAATACGATTCATGCAAGGGCT	1560
Oy	1641	ACATCAGCTGGGACTTCTTTTGGTGTACTGTGGGAAGTACGGGAACCAAGTGTAGAGC	17000
Db	1561	ACATCAGCTGGGACTTCTTTTGGTGTACTGTGGGAAGTACGGGAACCAAGTGTAGAGC	16200

QY	1701	ACACCCCTGGGACGGGTCCACACCCACAGGGGCCACTTCACAGGTGATCTGGATCTGATCTACGAC	1760
Db	1621	ACACCCCTGGGACGGGTCCACACCCACAGGGGCCACTTCACAGGTGATCTGGATCTGATCTACGAC	1680
QY	1761	GACTGGAGAACTGGGTCTGGGCGGAGATATGGTCTTTTGCCCATGGCTCTGGCCCTGGA	1820
Db	1681	GACTGGAGAACTGGGTCTGGGCGGAGATATGGTCTTTTGCCCATGGCTCTGGCCCTGGA	1740
QY	1821	GCCTTGAGCACACAGCTGCAGAGGCTGCAGGTGACCCGGAAAGCTGTCTGGAGATGGAGAGC	1880
Db	1741	GCCTTGAGCACACAGCTGCAGAGGCTGCAGGTGACCCGGAAAGCTGTCTGGAGATGGAGAGC	1800
QY	1881	AGGCGGCTTCTCTGTGGGAAAGCGCCACCCCTTGCTACTGTACTGTGGCCACAGCAACA	1940
Db	1801	AGGCGGCTTCTCTGTGGGAAAGCGCCACCCCTTGCTACTGTACTGTGGCCACAGCAACA	1860
QY	1941	GCAACAAGTGGGGTCAACCCCGGGGGCTACCGCATTCAGATGGTCAAGCTTGTGCTGGAGAGC	2000
Db	1861	GCAACAAGTGGGGTCAACCCCGGGGGCTACCGCATTCAGATGGTCAAGCTTGTGCTGGAGAGC	1920
QY	2001	CGCTGCCCAAAACAGCTCCATGGGCGAGAGGCTTCAGCTGGGAGAGATACAGCTGGGCTG	2060
Db	1921	CGCTGCCCAAAACAGCTCCATGGGCGAGAGGCTTCAGCTGGGAGAGATACAGCTGGGCTG	1980
QY	2061	TGACCCACAGCGGAGAGAGAGAGACCCAGTATGACAGACGCTTTTCAATAGATACCTT	2120
Db	1981	TGACCCACAGCGGAGAGAGAGAGACCCAGTATGACAGACGCTTTTCAATAGATACCTT	2040
QY	2121	GGGCGCCCACTGGGATTTAGTACGACTCATCAACAAATGAGACCAATTCCTGGAAAGGATT	2180
Db	2041	GGGCGCCCACTGGGATTTAGTACGACTCATCAACAAATGAGACCAATTCCTGGAAAGGATT	2100
QY	2181	TGTGGGCTGGGTACAGCTGTGTTTTCTGCATATCCACATGCAGAGACATTCCTTAACA	2240
Db	2101	TGTGGGCTGGGTACAGCTGTGTTTTCTGCATATCCACATGCAGAGACATTCCTTAACA	2160
QY	2241	CAGTGACTGTGGGAGACGCGGTGGGCTTCTCTCCGACCCATATACCTTTTGACGAG	2300
Db	2161	CAGTGACTGTGGGAGACGCGGTGGGCTTCTCTCCGACCCATATACCTTTTGACGAG	2220
QY	2301	ACCCTCCTTTACTTGGCCGAGCTCCATCTCCGAGGGAGACAGAGATGCTGGGGCT	2360
Db	2221	ACCCTCCTTTACTTGGCCGAGCTCCATCTCCGAGGGAGACAGAGATGCTGGGGCT	2280
QY	2361	GCAGAGTCAACCCCTAGCTTGGCTGGCCGAGGCTGCTGCTGTCGCCCGACCTCCCTG	2420
Db	2281	GCAGAGTCAACCCCTAGCTTGGCTGGCCGAGGCTGCTGCTGTCGCCCGACCTCCCTG	2340
QY	2421	CCTTCTCCACAGGGGCTTCTCTACACACTAGCGGCTCTGGGATGGGCACTGTGGCAA	2480
Db	2341	CCTTCTCCACAGGGGCTTCTCTACACACTAGCGGCTCTGGGATGGGCACTGTGGCAA	2400
QY	2481	GGGCGCCAGGGCCAGGGTGTGAGGAAATGGGAGACACTGGGAGTGGGCGGGCAGCGG	2540
Db	2401	GGGCGCCAGGGCCAGGGTGTGAGGAAATGGGAGACACTGGGAGTGGGCGGGCAGCGG	2460
QY	2541	TTTCCCTTTTCTGTGCCAGAGACTCTTTTCTTCCACTACCC	2581
Db	2461	TTTCCCTTTTCTGTGCCAGAGACTCTTTTCTTCCACTACCC	2501
RESULT 4			
ABK35481			
ID	ABK35481	standard; DNA; 2572 BP.	
XX	ABK35481;		
AC			
XX			
DT	08-MAY-2002	(first entry)	
XX			
DE	Human endometrial cancer related gene, AOC2.		
XX			
KM	Human; ds: gene: endometrial cancer; differential expression; DNA microarray; protein microarray.		

XX OS Homo sapiens.
XX PN MO200209573-A2.
XX PD 07-FEB-2002.
XX PE 31-JUL-2001; 2001WO-US24104.
XX PR 31-JUL-2000; 2000US-221735P.
XX PA (BGHM) BRIGHAM 6 WOMENS HOSPITAL INC.
XX PI Mutter GL;
XX DR WPI: 2002-179967/23.
XX PS P-PSDB; AAU84261.
PT Diagnosing endometrial cancer comprises determining expression of
PT nucleic acid molecules or expression products that are differentially
PT expressed in normal and malignant endometrium .
PS Claim 1; Page 37-38; 233pp; English.
XX The invention relates to diagnosing endometrial cancer in a subject
CC suspected of having endometrial cancer comprising determining the
CC expression of a set of nucleic acid molecules or expression products in
CC an endometrial sample suspected of being cancerous, where the set of
CC nucleic acid molecules comprises at least 2 nucleic acid molecules
CC selected from 50 fully defined sequences as given in the specification.
CC The nucleic acids are used as an array of at least 2 of the 50
CC nucleic acids bound to a solid substrate. Also included is a solid-phase
CC protein microarray comprising at least 2 antibodies or its antigen
CC binding fragments, that specifically bind at least 2 different
CC polypeptides from the 50 fully defined sequences as given in the
CC specification, fixed to a solid substrate. The methods and arrays are
CC useful for the diagnosis of endometrial cancer, selecting and monitoring
CC treatment regimes and identification of lead compounds useful for the
CC treatment of endometrial cancer. The present sequence is one of 50
CC genes differentially expressed between cancerous and non-cancerous
CC samples.
SQ Sequence 2572 BP; 538 A; 736 C; 669 G; 629 T; 0 other:
Query Match 25.2%; Score 1017.8; DB 24; Length 2572;
Best Local Similarity 67.6%; Pred. No. 9.7e-252;
Matches 1545; Conservative 0; Mismatches 637; Indels 105; Gaps 4;
QY 161 ATGACCAAGACAGCAATCTCGTCTCATTTCTGGCGGTATCCATCCATCTTTGGCTTG 220
DB 27 ATGCACTCTCAAGATATCTCTGGGCTTCTGCGACTGTCCCTCATTAACATCTTGGCCTG 86
QY 221 GTTGTGTCCTCTGTGTGGGAGGGGTGAGATGGGGGTGAACCCAGAGTTCCTCCAT 280
DB 87 GCGTATGTTTGTCTGACAGAGCCAGGTGT-----CGAGCAGGCTTCCAC 134
QY 281 TCCCTCTGTATCTCCAGTGCAGGCTTGGACACACCTGGCCAGAGGAGCTGTTT 340
DB 135 TCCCTCTGTATGCTATGCGATGGGCCAGCCCTGACCACCTGGCCAGAGGAGCTGTTT 194
QY 341 GGAGACTGAGCGAGAGAGAGCTGAGCGGTGATGCGGCTTGTGACCCAGGCGGTGGG 400
DB 195 GAGAGCTGAGCGAGAGAGAGTGTGACGTGTGATGCGTTCCTGACCCAGGCGGTGGG 254
QY 401 CCAGGCGTGTGATGACAGCCAGGCGGCTTGGACACACTGTCTTCTAGTGGAG 460
DB 255 CCAGGCGTGTGAGACGACCCAGGCTGACCTTGGACAACTGTCTTCTAGTGGAG 314
QY 461 TTGCACTGCTCCCAAGGCTGAGCGCTGCTCACTTGGAGAGGGGAGCCGCCACCT 520
DB 315 CTGACAGCTCCCGCCAAAGCTGAGCCCTGGCCACCTGGACAGGGGAGCCGCCACCT 374
QY 521 GCCCGGAGGAGCACTGGCATCGTCTTTTGGAGGCAACCCAGGCCAACTGAGTGA 580

DB 375 GCCCGGAGGAGCACTGGCATCGTCTTTTGGAGCAACCCAACTCATGTGATGAG 434
QY 581 CTGTGTGTGGGCGCACTGCTCACCCTCTCATATGCGGAGAGTGTGAGGCTCAT 640
DB 435 CTGTGTGTGGGCGCGCTGTCTCACCCTGTACATATGGGAGTGTGATGTGAGGCTC 494
QY 641 GGAGCGCCCTCTCCCTATACCGAGCGCCGCTGTCTTCAAGATGACTGACATAGAC 700
DB 495 GCGGCGCCCTCTCCCTATACCGCTGCGCGGTGTGAGAGTGTACACAGATGTGG 554
QY 701 CAGATGATTTCAACAGAGAGTGGCCCGGCTTCTGCGGCTTCTCACCAGTGTCTTC 760
DB 555 AGGCATCTGAAGATGTGGAGTACCCAAAGC-----ACCATCTTCTCTGTCCACC 608
QY 761 TACAAGCACCGGAGAGCAAGCTGTGACAAATGACACGCGTCCCGTGTGCAATCA 820
DB 609 TTCAACTCAATATGCTCTACCTGTGAGCTGTGACATGCCACCCCTGGGGCTTGGCTCA 668
QY 821 GGGAGCCGGCCCACTGTTGGCTTGTCTACTACAAACATTCGGGCGCTGGTTCTCTG 880
DB 669 AGGGAACGAATACCTGCGATTGGCCCTACATATCATCTCAGAGGCTGTCTTTCTT 728
QY 881 CACCACTGGGCTGTGAGCTGTAGTGAACCAAGGCGCTTGAACCCCTGCGGCTGACT 940
DB 729 CACCCCTGGGCTGTGAGCTGTAGTGAACCAAGGCGCTTGAACCCCTGCGGCTGACT 788
QY 941 ATCAAGAAAGTGTCTATCAAGCGCCTACTACAGACGCTGCGCCAGTGGAGCGCCAG 1000
DB 789 GTCCAGACAGCTTCTACTTGGGCACTATGACAGTCTGGGCGCATGGAGAGCGAG 848
QY 1001 TTTGAGGCGCGCTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1060
DB 849 TTTAATCTGTGGCGGTGTGAAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 908
QY 1061 TCCCTGAATCCCTCTGTGCGCCCGGCTGAGCTCCCTCTCAAGTGTCTATCCCAAGGC 1120
DB 909 TCCCTGAATCTGTGCAAGCTGTGCAAGCTGTGCAAGCTGTGCAAGCTGTGCAAGCT 968
QY 1121 CCCCGCTTGT 1180
DB 969 TCCAGTACATGCTGTGCAAGAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1028
QY 1181 CTGCGAGATTTAGTGTGCGCCCAAGATCTTGTGCTGTGTGTGTGTGTGTGTGTGT 1240
DB 1029 CATGGGCTGT 1088
QY 1241 TATGAGATTAAGCTTCAAGAGGCTTGGCATCTATGTGTGTGTGTGTGTGTGTGT 1300
DB 1089 TATGAGATTAAGCTTCAAGAGGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1148
QY 1301 ACAGCCGCTATGT 1360
DB 1149 CTGACTCGGTATTTGATATGACAGCTTGTGACTGTGCGCTGAACGCGAGGCTTGT 1208
QY 1361 GGGGTGACTGCGCTTACTTGTGCGCACTTGTGAGTGTGTGTGTGTGTGTGTGT 1420
DB 1209 GAGTGTGAGTGTGCGCTTATGAAGCGAGATGTGTGACATCATATATAGTGTGT 1268
QY 1421 GCCCCCAAGCAATAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1480
DB 1269 GAGTGTGAGTGTGCGCTTATGAAGCGAGATGTGTGACATCATATATAGTGTGT 1328
QY 1481 CGGCGACCACTCAATCTTCTACTGCACTATTTGGGGGTCTTGGGAAAGGTGTCTG 1540
DB 1329 CGAAGGACACAAATTAATCTTCAAAATCAATTTATAGTGTGTGTGTGTGTGTGT 1388
QY 1541 GTCTGAGATCTATGTCAACTTGTCAACTATGACTATGTGTGTGTGTGTGTGTGT 1600
DB 1389 GTGTGAGT 1448
QY 1601 CCCAGTGGGCGCATTAATATGATTTATGCAAGGCGTACATCATGTGTGTGTGTGT 1660


```

Db 1449 CCAATGGGACATTGAAGGCGGGTCCATCCACGGGTTATATCACACACGCTTCTG 1508
QY 1661 TTGTGCTACTGGG-----AGTAGGGGAACCAAGTGTGAGACACACCTTGGGACG 1714
Db 1509 AAAGGGGAGAGAGGGGCTCTCTTTGGGAACCGTGTGGGGGAAAGAGTGTCTGGAAACG 1568
QY 1715 GTCCACACCCACAGCGCCCACTTCAAGGTGATGTGATGAGCAGACGATGAGAGATGG 1774
Db 1569 GTGCACACACATGCTTCCACATTCAGCTGACCTGATGTGGCGAGGCTGAAAAAATG 1628
QY 1775 GTCTGGGCGGAGATGATGATGTTGTGTCCTGATGCTGCGCCCTGGAGCCCTGAGACAG 1834
Db 1629 GTGTAGCTGAAGACGATGATTTAAACCTGTGGCTGCCCCCTGGAAACCCGGAGACTGG 1688
QY 1835 CTGACAGAGCTGACGATGATGACCGGAAGCTGCTGAGATGAGAGACAGCCGCTTCTC 1894
Db 1689 CTACAGAGCCCAACGATGATGATGAGAGAGTCTGGGAAAGAGAGACCTACAGCTTTTCC 1748
QY 1895 GTGGGAAGCGGACCCCTGCTACTGATGATGCTGACCAACACACAGACAGATGGGGT 1954
Db 1749 TTGGGAAGCCCTTACCCCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1808
QY 1955 CACCCCGGCGCTACCGCATCCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2014
Db 1809 CACACGCGCGA----- 1820
QY 2015 AGCTCATGCGGAGAGGCTTACCTGAGTGAAGGATGACAGCTGCTGCTGCTGCTGCTG 2074
Db 1821 -----TACCACTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1847
QY 2075 GAGAGAGAGCCATGATGACAGAGCTTTCAATGACACCTTGGGCCCCCACTGTG 2134
Db 1848 GAGAGAGAGTACAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1907
QY 2135 GATTTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2194
Db 1908 ACCTTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1967
QY 2195 ACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2254
Db 1968 ACACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2027
QY 2255 AAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2314
Db 2028 AAGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2087
QY 2315 TCTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2374
Db 2088 TCCCTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2147
QY 2375 CTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2434
Db 2148 GTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2207
QY 2435 GCGTCT 2441
Db 2208 GCGTCT 2214

```

RESULT 5
 ABK63566
 ID ABK63566 standard; cDNA; 1051 BP.

XX Rat sequence differentially expressed in response to a hepatotoxin #1473.
 XX
 XX 18-JUN-2002 (first entry)
 XX
 DE Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
 KW differential expression; centrilobular necrosis; steatosis.
 OS Rattus norvegicus.

```

XX XX W0200210453-A2.
PN XX
PD XX
XX 07-FEB-2002.
XX 30-JUL-2001; 2001WO-US23872.
XX 31-JUL-2000; 2000US-222040P.
PR 02-NOV-2000; 2000US-244880P.
PR 11-MAY-2001; 2001US-290029P.
PR 15-MAY-2001; 2001US-290645P.
PR 22-MAY-2001; 2001US-292335P.
PR 06-JUN-2001; 2001US-295798P.
PR 13-JUN-2001; 2001US-297457P.
PR 19-JUN-2001; 2001US-298884P.
PR 09-JUL-2001; 2001US-303459P.
XX
XX (GENE-) GENE LOGIC INC.
XX Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;
XX WPI; 2002-241625/29.
XX
XX Predicting toxic effects of compounds or the progression of these toxic
XX effects by determining the changes in gene expression in tissues or
XX cells exposed to the toxin and comparing these to gene expression in
XX unexposed tissues or cells.
XX
XX Claim 1; Seq ID No 1473; 239pp; English.
XX
XX The invention relates to methods for predicting toxic effects of
XX compounds or the progression of these toxic effects by determining the
XX global changes in gene expression in tissues or cells exposed to the
XX toxin and comparing these to gene expression in unexposed tissues or
XX cells. Also included are methods of predicting at least one toxic
XX effect of a compound or progression of a toxic effect, preferably the
XX hepatotoxicity of a compound, comprising detecting the level of
XX expression in a tissue or cell sample exposed to the compound of two or
XX more genes listed in the specification, where differential expression of
XX the genes is indicative of at least one toxic effect or progression.
XX The method can also be used to identify an agent which modulates the
XX toxic response and predict cellular pathways that a compound modulates
XX in a cell. The methods utilise a set of at least two probes (on a solid
XX support in kit form), where each of the probes comprises a sequence that
XX specifically hybridises to a gene listed in the specification, a computer
XX system comprising a database containing information identifying the
XX expression level in a tissue or cell sample exposed to a hepatotoxin of a
XX set of genes comprising at least two genes listed in the specification,
XX and a user interface to view the information used to present information
XX identifying the expression level in a tissue or cell of at least one gene
XX listed in the specification. The method is useful for elucidating global
XX changes in gene expression and for identifying toxicity markers in
XX tissues or cells exposed to a known toxin. The genes may be used as
XX toxicity markers in drug screening and toxicity assays. The genes and
XX gene expression information may be used as diagnostic markers for the
XX prediction or identification of the physiological state of tissue or cell
XX sample that has been exposed to a compound or agent. Hepatotoxicity
XX is characterised by centrilobular necrosis and steatosis. The present
XX sequence is an expressed sequence tag (EST) or cDNA derived from a gene
XX which is differentially expressed in response to a hepatotoxic agent.
XX
XX Sequence 1051 BP; 200 A; 318 C; 291 G; 232 T; 10 other;

```

Query Match 17.7%; Score 714.4; DB 24; Length 1051;
 Best Local Similarity 82.3%; Pred. No. 1.1e-173;
 Matches 811; Conservative 5; Mismatches 169; Indels 0; Gaps 0;

```

QY 141 CTGTCCTCTTCGTGGGAAATGAACAGAGACAACTCTGCTCTCTCTCTGCGCG 200
Db 66 CTGTCACCTTCAGAGAAATGACCAACCACTGCTCTCTCTCTCTGCGCGGTG 125
QY 201 TCATCACCATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 260

```

```

Db 126 TCATCATCATCTTCCTTGGTTGTTGCTTCTAGTGTGACAGAGGAGATGGGGCA 185
OY 261 AACCCAGCAGCTTCCCATTTGCCCTCTGTATCTCCAGTCCCAAGCTTGGACACACC 320
Db 186 GACTGAGCCAACTTATTCCTTCCCTTCTTCTAGCCTCAGCCCAAGCCCAAGACT 245
OY 321 CTGGCCAGAGCAGCTTTTGCAGACTGAGCCGAGAGAGAGTGAAGGCTGATGAGCT 380
Db 246 CTGGCCAGAGCAGCAGCTTTGACAGACTGAGCCCTGAGAGAGTGAAGTGTGATGAGCT 305
OY 381 TTTCAGCCAGCGGCTGGGGCCAGGGCTGTGTGATGACAGCCAGCCCGGCTCGGACA 440
Db 306 TTCTGATCAAGACCTTGGGGCCAGGGCTGTGTGATGACAGCCAGCTGACCTCGGACA 365
OY 441 ACTGTCTCTCATGTGAGTGTGACAGTCCCTCCCAAGGCTGACGCTGAGCTCACTTGG 500
Db 366 ACTGTCTCTCATGTGAGTGTGACAGTCCCTCCCAAGGCTGACGCTGAGCTGAGCTG 425
OY 501 ACAGGGGAGAGCCCGCCAGCTGCGGGAGAGCACTGGCCATCGCTTCTTGGCAGGACAC 560
Db 426 ACAGAGGGGGGGCCCCAGCCCGTGGGGAGAGCACTGGCCATCATCTTCTTGGTGAACAC 485
OY 561 CCCAGCCCAAGCTGAGTGTGAGTGTGGGGCCACTGCTTCACTCCCTCTCATATGCGGG 620
Db 486 CCAGACCTATGTGAGCAGTGTGTGGGGCCCGCCCTGCTTCACTCCCTCATATGCGGG 545
OY 621 ACCTGAGTGTGAGGCTATGAGAGGCCCCCTGCTTCACTCCAGAGCCCGCTGTCTTC 680
Db 546 AGTGTAGTGTGAGGCTGTGAGGCCCCCTGCTTCACTCCAGAGGCTGTCTGTACCA 605
OY 681 AAGATACCTGAGATAGACAGATATCTTCAACAGAGAGTGGCCCAAGCTTCTGGGAC 740
Db 606 GAGATATCAGAGATATTCAGAGAGATCTTTCACAGAGAGTGGCCCAAGCTGTCTGTC 665
OY 741 TTCTCACCACCTGTGTCTTCAACAGACCGGGAGAGGAACTGTGACATATGACCCAG 800
Db 666 TCCTCATCTACTGTGTCTTCAACAGACCGGAAAGCACTGTGTAAGGCTGTGACAT 725
OY 801 CTCCCGTGTGTGATATGAGGGAGCGGGGCACTGCTGTGCTTCTCAACATATCT 860
Db 726 CCCCCGTGTGTGATATGAGGGAGCGGGGCACTGCTGTGCTTCTCAACATATCT 785
OY 861 CGGGCGTGTGTCTTCTTCTGACACAGTGGGCTGTGAGCTGTGTAAGCAACAGGCCC 920
Db 786 CAGGGGCTGTGTACCTTACCCATGCGCTTAGAGCTTGTGTGATCACAAGGCC 845
OY 921 TTGACCTGCGCGCTGTGACTATCCAGAGGTTCTATCAGAGCGCTACTAGACAGCC 980
Db 846 TGGATCTGTGCTGTGTGACATCCAGAAAGTATTCACCAAGGCGCTACTATGAGAGTC 905
OY 981 TGGCCAGTGTGAGGCGCCAGTTGAGCGCGGCTGTGTAATGTGTGCTGATCCCGACA 1040
Db 906 TGACAGTGTGAGGAGCATGTGTGAGGCTGTGTAATGTGTGCTGATCCCGACA 965
OY 1041 ATGGCAGAGTGTGTCTGTGCTGTGAGTCCCTGTGCTGAGGCTGAGCTCCCGCTC 1100
Db 966 ATGGCAGAGTGTGTGCTGTGCTGTGAGTCCCTGTGCTGAGGCTGAGCTCCCGCTC 1025
OY 1101 TACAGTGTATCCCAAGGCCCCCG 1125
Db 1026 TGCARTTCAYCCNGARGNCNMG 1050

```

RESULT 6
ABL62445

ID ABL62445 standard; DNA: 2441 BP.

XX ABL62445;

XX 15-MAY-2002 (first entry)

DE Colon adenocarcinoma related gene sequence SEQ ID NO:782.

XX

```

KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumor; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;
KW gene; ds.
OS Homo sapiens.
PN WO200194629-A2.
XX 13-DEC-2001.
PD 30-MAY-2001; 2001WO-US10838.
PF 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 28-SEP-2000; 2000US-236412P.
PR 29-SEP-2000; 2000US-236891P.
PR 29-SEP-2000; 2000US-236892P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
PA (AVAL-) AVALON PHARM.
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrihan S;
PI Soppet DR, Weaver Z;
XX WPI: 2002-188264/24.
XX Screening for anti-neoplastic agent involves exposing cells to a
XX chemical agent to be tested for anti-neoplastic activity, and
XX determining a change in expression of a gene of a signature gene set
XX
XX Claim 1; SEQ ID 782; 44pp; English.
XX The present invention describes a method (M1) for screening for an
XX anti-neoplastic agent. The method involves exposing cells to a chemical
XX agent to be tested for anti-neoplastic activity, determining a change in
XX expression of at least one gene (I) of a signature gene set, where (I)

```


Db 2014 ACCAGACAGACCCCTGGACACCCCGCTGTCTTTGAGACGTTTCTTACACAGAGAGA 2073
 QY 2163 CCATTCTCTGMAAGATTGGTGGCTGGTGACACTGGTTTTCGATATCCACATG 2222
 Db 2074 ACATTGMAATGAGGACCTGGTGGCTGGTGACAGGCTTCCGACATCCCCACT 2133
 QY 2223 CAGAGACATTTCTTACACAGTACTGTGGGAAAGCGGCTTCTTCCCGAGCCCT 2282
 Db 2134 CAGAGACATTTCCCAACAGCCACACTGGGAACTCCGTTCTGCTTCCGGGCAT 2193
 QY 2283 ATACTCTTTTGACGAGACCCCTCTTACTCTGCGACTC 2325
 Db 2194 TCACCTTTTCCAGAGACCCCTCTCTGCAATCCAGAGACAC 2236
 RESULT 7
 ABL67675
 ID ABL67675 standard; DNA; 2441 BP.
 XX
 AC ABL67675;
 XX
 DT 15-MAY-2002 (first entry)
 XX
 DE Oesophagus cancer related gene sequence SEQ ID NO:6012.
 XX
 KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
 KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200194629-A2.
 XX
 PD 13-DEC-2001.
 XX
 PF 30-MAY-2001; 2001MC-US10838.
 XX
 PR 05-JUN-2000; 2000US-209473P.
 PR 05-JUN-2000; 2000US-209531P.
 PR 18-SEP-2000; 2000US-23313P.
 PR 18-SEP-2000; 2000US-233617P.
 PR 20-SEP-2000; 2000US-234009P.
 PR 20-SEP-2000; 2000US-234034P.
 PR 20-SEP-2000; 2000US-234052P.
 PR 22-SEP-2000; 2000US-234509P.
 PR 22-SEP-2000; 2000US-234567P.
 PR 25-SEP-2000; 2000US-234923P.
 PR 25-SEP-2000; 2000US-234924P.
 PR 25-SEP-2000; 2000US-235077P.
 PR 25-SEP-2000; 2000US-235082P.
 PR 25-SEP-2000; 2000US-235134P.
 PR 25-SEP-2000; 2000US-235280P.
 PR 26-SEP-2000; 2000US-235637P.
 PR 26-SEP-2000; 2000US-235638P.
 PR 27-SEP-2000; 2000US-235711P.
 PR 27-SEP-2000; 2000US-235720P.
 PR 27-SEP-2000; 2000US-235840P.
 PR 27-SEP-2000; 2000US-235863P.
 PR 28-SEP-2000; 2000US-236028P.
 PR 28-SEP-2000; 2000US-236032P.
 PR 28-SEP-2000; 2000US-236033P.
 PR 28-SEP-2000; 2000US-236034P.
 PR 28-SEP-2000; 2000US-236109P.
 PR 28-SEP-2000; 2000US-236111P.
 PR 29-SEP-2000; 2000US-236842P.
 PR 29-SEP-2000; 2000US-236891P.
 PR 29-SEP-2000; 2000US-237172P.
 PR 02-OCT-2000; 2000US-237173P.
 PR 02-OCT-2000; 2000US-237278P.
 PR 02-OCT-2000; 2000US-237294P.
 PR 02-OCT-2000; 2000US-237295P.
 PR 02-OCT-2000; 2000US-237316P.

PR 03-OCT-2000; 2000US-237425P.
 PR 03-OCT-2000; 2000US-237598P.
 PR 03-OCT-2000; 2000US-237604P.
 PR 03-OCT-2000; 2000US-237606P.
 PR 03-OCT-2000; 2000US-237608P.
 PR 01-NOV-2000; 2000US-244867P.
 PR 01-NOV-2000; 2000US-245084P.
 XX
 PA (AVAL-) AVALON PHARM.
 PI Young PE, Augustus M, Carter KC, Edner R, Endress G, Horrihan S;
 PI Soppet DR, Weaver Z;
 XX WPI; 2002-188264/24.
 DR
 XX
 PT Screening for anti-neoplastic agent involves exposing cells to a
 PT chemical agent to be tested for anti-neoplastic activity, and
 PT determining a change in expression of a gene of a signature gene set -
 PS Claim 1; SEQ ID 6012; 44pp; English.
 XX
 XX The present invention describes a method (M1) for screening for an
 CC anti-neoplastic agent. The method involves exposing cells to a chemical
 CC agent to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening
 CC an anti-neoplastic agent, and can be used for producing a product which
 CC is the data collected with respect to the anti-neoplastic agent as a
 CC result of M1, and the data is sufficient to convey the chemical
 CC structure and/or properties of the agent. M1 can be used in the
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
 CC carcinoma, papillary carcinoma and Wilm's tumour.
 XX
 XX Sequence 2441 BP; 515 A; 789 C; 684 G; 453 T; 0 other;
 SQ
 Query Match 10.5%; Score 422.2; DB 24; Length 2441;
 Best Local Similarity 52.8%; Pred. No. 4.3e-98;
 Matches 1099; Conservative 0; Mismatches 918; Indels 66; Gaps 6;
 QY 303 CCCAGCCTTGACACACCCCTGGCCAGACGCTGTTGACAGACTGACCCGAGAGAGC 362
 Db 160 CCTCCCGGGGAGTCTGCCAGGAAGGAGGGGTGTTTTCAGACTTAAGCAACCAAGAGC 219
 QY 363 TGACGGCTGTGATCGCTTCTGTACCCAGGGGCTGGGGCCAGGGCTGTGATCAGCCC 422
 Db 220 TGAAGGCGATGCACAGCTTCTGTGTCAAGAGGAGCTGAGCTCAGCCCTCCAGTA 279
 QY 423 AGGCGCGCCCTGAGCAACTGTCTTTCAGTGAAGTGCACCTGCTCCCAAGGCTG 482
 Db 280 CCACACCATAGGCAAGAACACCGTGTTCATCGAGATCTCTGCCAAGAAAGTACC 339
 QY 483 CAGCCCTGCTCACTTGAGCAGAGGGGAGCCGCCACTGCCCGGAGGACTGGCCATCG 542
 Db 340 ATGTGCTGAGTTCGTGATTAAGTGAAGAGCATCTGTGCGGGAAGCCGTCGCTCA 399
 QY 543 TCTTCTTTGGCAGGACCCAGCCCAACGATGATGAGCTGTGTGGGGCATGTGCTC 602
 Db 400 TCTTCTTTGGCAGGAGAGCATCCCAATGACAGGTTGCTGTGTGGGGCCCTGCAG 459
 QY 603 ACCCTCTCATGCTGGGAGCTGACTGTGAGC--GTCATGAGAGCCCTGCTATC 659
 Db 460 GGGCTCTCATGCTGAGGAGCTGTCCCGGAGGCTGGGTACCATCTCTCTGGGCATCGA 519
 QY 660 ACCGAGCCCGCTGCTGTTCCAGAGTACCTGACATAGACAGATGATCTTCAACAGAG 719
 Db 520 GGCCATCTCCACAGAGAGATATGCCCTCTTACACACACCTTCAGAGAACCCACCAAGC 579

QY 720 AGTGGCCCAAGGCTTGGGCTTCTCCACACTGTGCTTCTACAGACCGGGACGGA 779
 Db 580 CCGGCGATCATCTTCTCTCAATATACACAGAGCTTCTCAATTCAGAACTGATGACAT 639
 QY 780 ACCTGTGACAAATGACACAGGCTCCCGTGGTGTGCAATACAGGGACCGGCACTGGT 839
 Db 640 GCGTGGCTTACAGATGTGGCCCGCGGGGTGTGGCTTCTGGCCAGCGCCGCACTTGGC 699
 QY 840 TTGGCTCTACTACACACTCTCGGGCGCTGGGCTTCTCTGACACAGCTGGGCTTGAAGC 899
 Db 700 TTATCATACAGCGCTATGTAGAAAGC-----TACTTCTGACACCCCACTGGGCTGGAGC 753
 QY 900 TGGTAATGAACACAGGCGCTTACCCCTGCGGCTGACATATCCAGAAAGTGTCTATC 959
 Db 754 TCTGTGTGATCATAGGAGACACATGTGTGGCACTGGGCGGTGGAGAGAGTGTGTCA 813
 QY 960 AAGGCGGCTACTACAGACAGCTGGCCAGCTGAGAGGCGCAATTTGAGGCGCGGCTGTGA 1019
 Db 814 ACGGGAAGTTCTATGTGGAGCCCAAGAACTGGCTCGGAATATGCAAGATGAGAGGTGG 873
 QY 1020 ATGTGTGCTGATCCAGACA-----ATGGCAACAGGTGGGT 1055
 Db 874 ACGTGTGTGCTGTGGAGACCGGCTGCTGGGGCAAGGGGCAATGACAGACAGAGAGC 933
 QY 1056 CCTGTGCTGAAAGTCCCGCTGTGGCCCGGGTCCAGCTCC----- 1095
 Db 934 GCGCCCTCTTCTCTCCACAGCGCCGCGGGGACTCCCAAGCCCAATCATGTGACG 993
 QY 1096 -CCCTTCACTTATATCCCAAGGCGCCGCTTCACTGTGACAGAGAGTGAAGTGGCT 1154
 Db 994 GCGCCCGTGTGTGACGCGCCAGCGCCCTGCTTCAAGGCTGAGAGGCAACGCTGTGCT 1053
 QY 1155 CCTCACTGTGACTTTCCTTGTGGCTGTGAGCATTCAGTGGCCCAAGATCTTTGACG 1214
 Db 1054 ACGGCGGTGTGAGGCTTTCCTTCGCGGCTGCTCTCTTGGGGGTGAGGCTGGAAGC 1113
 QY 1215 TTGCTTCCAGAGAAAGACTATGATATGAGATTAAGCTCCAAAGGCGCTTGGCACT 1274
 Db 1114 TGCACCTTGGGGGAGAGGCAATGCTTATGAGTCAAGCTGCAAGAGCACTGGGCTGT 1173
 QY 1275 ATGTGAAATTCGCCAGACCAATGACAGACCGCTATGTGATGAGAGCTTTGGCATG 1334
 Db 1174 ATGGAGACACACACACTGACAGCATGACAGACAGTACTGATGTGGGCGGCGCTGG 1233
 QY 1335 GCAAGTACACACGCGCCGTGACCGGTGGGTGAGTGGCCCTACTTGGCCACTACGTGG 1394
 Db 1234 GCAGGCTCACTCATGATTAGCCCGCCGATGACTGCGCGAGAGACCGCACTTCTGTG 1293
 QY 1395 ACTGGCACTTCTTTGGAGTCCAGAGGCCGCCAAGACAAATACGTATGCTTTGTGT 1454
 Db 1294 AACTTTCCTACTATATATGATGCGCATGACCGGTCAATATCCCGAGCCCTGCTCT 1353
 QY 1455 TTGAACAGAACAGGCGCTCCCGTGGGGAGACAC-----ACTCAGATCTCTACT 1505
 Db 1354 TTGAAATGCCACAGAGGTGCCCTTGTGGCGGCACTTAATTCACACTTTAAAGTGGCT 1413
 QY 1506 GCGACTACTTGGGGGTCTTGGGAAACGATGCTGTGTGATGATCTATGTCCACTTGC 1565
 Db 1414 TCACTTCTATGCGGGCTGAAGGGCAGGTGTGTGCTGTGGGCACTTCACTGTCT 1473
 QY 1566 TCACTATGACTATGTGGGATACGCTTCCACACCCAGTGGGGCACTAGAAATACAT 1625
 Db 1474 ACAATTAATGATTACTTTGGGACTTATCTTACCCCAAGGGGTGATGGAGGCAAGA 1533
 QY 1626 TCTATGCCAGGCGTACATCAGCTGCGATTCCTTGTGTGCTACTGGAAATACGGGA 1685
 Db 1534 TGCATGCCACTGGCTACGTCACGCACTTCTACACCCCGAGGGGCTGCGCCACGGCA 1593
 QY 1686 ACAAGTGTAGAGACACCTTGGGACGCTTCACACCCAGCGCCCACTTCAAGGTGG 1745
 Db 1594 CTGCGCTGTACACACCTGATTGGCAACATACACTACTTGTGTGACACTGCGGTAG 1653
 QY 1746 ATCTGATGTAGCAGACTGGGAACTGGGTCTGGCGGAGGATATGGTCTTGTGCCCA 1805

Db 1654 ACTGTGATGTGGACGACCAAGAAAGCTTCCAGACACTGCAATGAAGTAAAGTAAACA 1713
 QY 1806 TGGCTGTGCCCTGGAGCCCTGAGCAGCACTGTAGAGGCTGACAGTACCCGGAAGTGC 1865
 Db 1714 TCACCAACCCCTGGAGCCCAAGACACCGGCTGTGTCCAGCAACTGTGAGAGACGAGT 1773
 QY 1866 TGGAGATGAGAGACAGCGCCCTTCTCTGTGGAAAGCGCACCCCTGCTACTGTATCC 1925
 Db 1774 ACTCTGGAGACGCGCAGCGGCTTCCGCTTCAAAAGAAAGCTGCTAAGTATGCTGTCT 1833
 QY 1926 TGGCAGCAACACAGACAAAGTGGGTCAACCCCGGGGTACCGGATCCAGATGTCTCA 1985
 Db 1834 TTACAGCCCCACAGAGAAACCTTGGGGCCACAAAGCGATACCGCTCGCATCACT 1893
 QY 1986 GCTTGTGTGAGAGCCGCTGCGCCCAAAACAGCTCATATGAGAGGCTTACGTGGAGA 2045
 Db 1894 CCATGTGCGGACAGAGTGTGTGCGCCCAAGCTGTGACAGAGAGACAGCCATCACTGGGCA 1953
 QY 2046 GGTACCAAGCTGTGTGACCCAGCGGAAAGAGAGAGCCAGTACAGACAGCGTTTCA 2105
 Db 1954 GGTACCCCTGTGAGTACCAAGTACCGGAGTGTGAGCTGTGAGAGCAGCATCTTACC 2013
 QY 2106 ATCAAAATGAGCTTGGGCGCCCACTGTGATTCAGTCACTTCAAT---CAACAATGAGA 2152
 Db 2014 ACCAAGACAGCCCTGTGACCGCGCGGTGTGTGAGCACTTCTTCAACAACAGAGA 2073
 QY 2163 CCATTGTGGAAGATTTGTGTGCTGTGAGACAGTGTGTTTGTGATATCCACATG 2222
 Db 2074 ACATTGAAATAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2133
 QY 2223 CAGAGGACATTTCTAACACATGACTGTGGGAAAGGCGTGGGCTTCTCTCGACCT 2282
 Db 2134 CAGAGGACATTTCTAACACACACACACCTGTGGAACTCCGTGGGCTTCTCTCGGCAAT 2193
 QY 2283 ATAACTTCTTGAAGAAAGCCCTTCTTCACTGTGCGGCACTC 2325
 Db 2194 TCACTTCTTCCAGAGAGACCCCTTCTGCGATCAGAGACAC 2236

RESULT 8
 AAC04285
 ID AAC04285 standard; cDNA: 401 BP.
 XX AAC04285;
 AC 06-OCT-2000 (first entry)
 AC 06-OCT-2000 (first entry)
 DE Human secreted protein 5' EST, SEQ ID NO: 8360.
 DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 KW Homo sapiens.
 OS Homo sapiens.
 PN EP1033401-A2.
 PD 06-SEP-2000.
 PF 21-FEB-2000; 2000EP-0200610.
 PR 26-FEB-1999; 99US-0122487.
 PA (BEST) GENSET.
 PA Dunas Milne Edwards J, Duclert A, Giordano J;
 DR WPI: 2000-500381/45.
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX

PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
PI Dmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG04942.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 4933; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIGO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1545 BP; 372 A; 395 C; 429 G; 349 T; 0 other;
Query Match 7.4%; Score 297; DB 23; Length 1545;
Best Local Similarity 98.4%; Pred. No. 6.7e-66;
Matches 300; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1457 GAACAGAACGAGGCGCTCCCTCGGGGAGACACCTCAGATCTTACGCACTACTT 1516
DB 931 GAGATGAACGAGGCGCTCCCTCGGGGAGACACCTCAGATCTTACGCACTACTT 990
QY 1517 GGGGCTTGGCGAAACGCTGCTGCTCAGATCTATGTCACCTTGCCTCAATATGAC 1576
DB 991 GGGGCTTGGCGAAACGCTGCTGCTCAGATCTATGTCACCTTGCCTCAATATGAC 1050
QY 1577 TATGTGTGGGATACGCTTTCACCCAGTGGGGCATAGAAATTCATTCATGCCACG 1636
DB 1051 TATGTGTGGGATACGCTTTCACCCAGTGGGGCATAGAAATTCATTCATGCCACG 1110
QY 1637 GAGTACATCAGCTCGGCACTTCTTTGTGTACTGGGAAGTAGGGAGCAAGTGCA 1696
DB 1111 GGTATACATCAGCTCGGCACTTCTTTGTGTACTGGGAAGTAGGGAGCAAGTGCA 1170
QY 1697 GAGCAGACCTGGGAGCGTCCACACCCAGCCCACTTCAGATGGATGTGATGTA 1756
DB 1171 GAGCAGACCTGGGAGCGTCCACACCCAGCCCACTTCAGATGGATGTGATGTA 1230
QY 1757 GCAGG 1761
DB 1231 GCAGG 1235
RESULT 11

ABQ54593
ID ABQ54593 standard; CDNA: 747 BP.
XX
AC ABQ54593:
XX
DT 22-AUG-2002 (first entry)
XX
DE Human ovarian antigen HVBK73 CDNA, SEQ ID NO:473.
XX
KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antiinflammatory; gynaecological; reproductive; chromosome 7q34-ter;
KW gene; ss.
XX
OS Homo sapiens.
XX
PN WO200200677-A1.
XX
PD 03-JAN-2002.
XX
PF 07-JUN-2001; 2001WO-US18569.
XX
PR 07-JUN-2000; 2000US-209467P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
DR WPI: 2002-147878/19.
DR P-PSDB; ABP41516.
XX
PT Isolated nucleic acid molecules, encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g.
PT ovarian cancer), immune disorders, cardiovascular disorders and
PT neurological diseases -
XX
PS Claim 1; SEQ ID NO 473; 2922pp; English.
XX
CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vulvinitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents cDNA encoding a human ovarian antigen of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed


```

Db      169  GCCAGGCGCCCTTCCTTCAAAAAGGAGAGCTGCTTAAGTACCTGCTTTTACCGACGCCCC 110
Qy      1938 ACAGCAACAAGTGGGGGTACACCCCGGGGCTACCGCATCAGATGCTCAGTTTCTGGAG 1997
Db      109  AGGAGAACCCCTGGGGGACACAGCGACGTAACGCCCTGCAGATCCATCTCATGGCGCACC 50
Qy      1998 AGCGGCTGCCCCCAAAACAGTCTCCATGGCGAGAGGCTTCACTGAGTGGAGAG 2046
Db      49  AGGTGCTGCCCCCAGCGCTGGCAGAGGAGGACGACATCACTGCGGCAAG 1

RESULT 15
ABA31594/c
ID      ABA31594 standard; DNA; 289 BP.
XX
AC      ABA31594;
XX
DT      23-JAN-2002 (first entry)
XX
DE      Probe #10060 for gene expression analysis in human heart cell sample.
XX
KM      Human: gene expression; heart; microarray; vascular system; probe;
KM      cardiovascular disease; hypertension; cardiac arrhythmia;
KM      congenital heart disease; ss.
XX
OS      Homo sapiens.
XX
PN      WO200157274-A2.
XX
PD      09-AUG-2001.
XX
PF      30-JAN-2001; 2001WO-US00666.
XX
PR      04-FEB-2000; 2000US-0180312.
PR      26-MAY-2000; 2000US-0207456.
PR      30-JUN-2000; 2000US-0608408.
PR      03-AUG-2000; 2000US-0632366.
PR      21-SEP-2000; 2000US-0234687.
PR      27-SEP-2000; 2000US-0236359.
PR      04-OCT-2000; 2000GB-0024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR      WPI; 2001-488899/53.
XX
PT      Single exon nucleic acid probes for analyzing gene expression in human
PT      hearts -
XX
PS      Claim 4; SEQ ID No 10060; 530pp; English.
XX
CC      The present invention relates to single exon nucleic acid probes for
CC      measuring human gene expression in a sample derived from human heart. The
CC      present sequence is one such probe. The probes may be used for
CC      predicting, measuring and displaying gene expression in samples derived
CC      from the human heart via microarrays. By measuring gene expression, the
CC      probes are useful for predicting, diagnosing, grading, staging,
CC      monitoring and prognosing diseases of the human heart and vascular system
CC      e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC      congenital heart disease.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pcr_sequences.
XX
SQ      Sequence 289 BP; 37 A; 77 C; 102 G; 73 T; 0 other;

Query Match      1.7%; Score 69.8; DB 22; Length 289;
Best Local Similarity 52.6%; Pred. No. 1.1e-07;
Matches 152; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
Qy      1758 CAGGACTGAGAACTGGGTGCGGCGGAGATATGCTTGTGCTTGTCCCATGGCTGTGCCCT 1817

```

```

Db      289  CAGGCAACCAAGAACAGCTTCCAGACACTGCAGATGAAGTAAGAAACATACCAACCCCT 230
Qy      1818 GGAGCCCTGAGACACACAGCTGCAGAGCTGCAGTGACCGGAAGCTGCTGAGATGGAG 1877
Db      229  GGAGCCCAACACACCGGCTGTGTCCAGCACTGTGAGCAGAGCAGTACTCTCGGGAGC 170
Qy      1878 AGCAGGCCGCTTCTCTGTGGGAAGCGCCACCCCTGCTTACCTGTACCTGCGCAGCAAC 1937
Db      169  GCCAGGCGGCTTCCTCCGCTTCAAAAAGGAGAGCTGCTTAAGTACCTGCTTTTACCGACGCCCC 110
Qy      1938 ACAGCAACAAGTGGGGGTACACCCCGGGGCTACCGCATCAGATGCTCAGTTTCTGGAG 1997
Db      109  AGGAGAACCCCTGGGGGACACAGCGACGTAACGCCCTGCAGATCCATCTCATGGCGCACC 50
Qy      1998 AGCGGCTGCCCCCAAAACAGTCTCCATGGCGAGAGGCTTCACTGAGTGGAGAG 2046
Db      49  AGGTGCTGCCCCCAGCGCTGGCAGAGGAGGACGACATCACTGCGGCAAG 1

```

Search completed: May 20, 2003, 03:47:38
Job time : 841 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2003, 03:31:13 ; Search time 214 Seconds
(without alignments)
5789.599 Million cell updates/sec

Title: US-10-081-408-1

Perfect score: 4040
Sequence: 1 gtcctccacccttagtc.....cccaaaaaaaaaaaaaa 4040

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 862724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/2/lna/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/lna/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/lna/5A_COMB.seq: *
4: /cgn2_6/ptodata/2/lna/5B_COMB.seq: *
5: /cgn2_6/ptodata/2/lna/PTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/lna/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85.8	2.1	7218	1	US-08-232-463-14
2	51.6	1.3	4386	4	US-09-300-0088-1
3	49	1.2	289	4	US-09-007-005-17
4	49	1.2	289	4	US-09-244-796-17
5	46.4	1.1	1571	4	US-09-276-531-49
6	45	1.1	2061	2	US-08-960-022-11
7	44.2	1.1	421	1	US-08-480-784-24
8	44.2	1.1	421	1	US-08-483-553-24
9	44.2	1.1	421	1	US-08-487-002-24
10	44.2	1.1	421	1	US-08-483-554B-24
11	44.2	1.1	421	1	US-08-488-011B-24
12	44.2	1.1	421	1	US-08-850-727-24
13	44.2	1.1	421	4	PCT-US95-10202-24
14	44.2	1.1	421	5	PCT-US95-10203-24
15	44.2	1.1	421	5	PCT-US95-10220-24
16	43.6	1.1	7218	1	US-08-232-463-14
17	43	1.1	7210	2	US-08-257-863B-10
18	43	1.1	7210	4	US-08-367-841A-10
19	43	1.1	7210	5	PCT-US95-07201-10
20	43	1.1	14581	4	US-08-520-373D-4
21	43	1.1	22481	4	US-08-367-841A-43
22	43	1.1	22481	5	PCT-US95-07201-43
23	43	1.1	22484	4	US-09-875-223-2
24	42.4	1.0	53526	3	US-08-658-136-2
25	42.4	1.0	53577	3	US-08-658-136-1
26	42.4	1.0	4403765	4	US-09-103-840A-2
27	42.4	1.0	4411529	4	US-09-103-840A-1

c 28	42	1.0	12001	1	US-08-458-568A-11	Sequence 11, Appl
c 29	41.6	1.0	90050	4	US-09-245-041-5	Sequence 5, Appl
c 30	41	1.0	51259	3	US-08-781-891-209	Sequence 209, App
c 31	40.8	1.0	87350	3	US-08-781-891-79	Sequence 79, Appl
c 32	40.8	1.0	87543	4	US-09-791-211-3	Sequence 3, Appl
c 33	40.6	1.0	32042	4	US-09-245-281-44	Sequence 44, Appl
c 34	40.4	1.0	43795	3	US-08-742-185-101	Sequence 101, App
c 35	40	1.0	112132	4	US-09-741-150-3	Sequence 3, Appl
c 36	39.8	1.0	789	4	US-09-149-476-288	Sequence 288, App
c 37	39.8	1.0	3381	4	US-09-009-119-1	Sequence 1, Appl
c 38	39.8	1.0	3381	4	US-09-371-507-1	Sequence 1, Appl
c 39	39.8	1.0	3383	5	PCT-US95-09098-1	Sequence 1, Appl
c 40	39.8	1.0	3404	1	US-08-265-429A-1	Sequence 1, Appl
c 41	39.8	1.0	3404	5	PCT-US95-09069-1	Sequence 1, Appl
c 42	39.6	1.0	1056	4	US-09-524-162-1	Sequence 1, Appl
c 43	39.6	1.0	1335	2	US-08-985-090-3	Sequence 3, Appl
c 44	39.6	1.0	1335	2	US-09-165-543-3	Sequence 3, Appl
c 45	39.6	1.0	1335	3	US-09-167-354-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOAMPOX VIRUS
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; City: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMTU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ9pl-F15
; US-08-232-463-14
Query Match 2.1%; Score 85.8; DB 1; Length 7218;

NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347-09
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 421 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-488-011B-24

Query Match 1.1%; Score 44.2; DB 1; Length 421;
Best Local Similarity 50.5%; Pred. No. 0.11;
Matches 100; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 2870 TGGCTGTCACATCCGCTCCATCCAGGCTCTTCTGCTCTCTCTCTAC 2929
DB 117 TCTTNCCTTCTCTTCCCTCCCTCCCTCCCTCTCTCTCTCTCTCTCT 176
QY 2930 CTACTTCT 2989
DB 177 GTCTACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 236
QY 2990 CCTGAGGAGATATCCATATCCAGCCCTGTACTCCCGCAGCCTCATGTC 3049
DB 237 CCTTCT 296
QY 3050 AAGTCCGCTCTCTCTCC 3067
DB 297 CAGTGGCGTATCTCGNC 314

RESULT 12
US-08-850-727-24
Sequence 24, Application US/08850727

Patent No. 6162897

GENERAL INFORMATION:

APPLICANT: Skolnick, Mark H.

APPLICANT: Goldgar, David E.

APPLICANT: Miki, Yoshio

APPLICANT: Swenson, Jeff

APPLICANT: Kamb, Alexander

APPLICANT: Harshman, Keith D.

APPLICANT: Shattuck-Eidens, Donna M.

APPLICANT: Tavtigian, Sean W.

APPLICANT: Wiseman, Roger W.

APPLICANT: Futreal, P. Andrew

TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer

TITLE OF INVENTION: Susceptibility Gene

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP

STREET: 1201 New York Avenue, N.W., Suite 1000

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/850,727

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/483,554

FILING DATE: 07-JUN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/348,824

FILING DATE: 29-NOV-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/308,104

FILING DATE: 16-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/300,266

FILING DATE: 02-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/289,221

FILING DATE: 12-AUG-1994

ATTORNEY/AGENT INFORMATION:

NAME: Ihnen, Jeffrey L.

REGISTRATION NUMBER: 28,957

REFERENCE/DOCKET NUMBER: 24884-109347

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-962-4810

TELEFAX: 202-962-8300

INFORMATION FOR SEQ ID NO: 24:

SEQUENCE CHARACTERISTICS:

LENGTH: 421 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

US-08-850-727-24

Query Match 1.1%; Score 44.2; DB 4; Length 421;
Best Local Similarity 50.5%; Pred. No. 0.11;
Matches 100; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 2870 TGGCTGTCACATCCGCTCCATCCAGGCTCTTCTGCTCTCTCTCTAC 2929
DB 117 TCTTNCCTTCTCTTCCCTCCCTCCCTCCCTCTCTCTCTCTCTCTCT 176
QY 2930 CTACTTCT 2989
DB 177 GTCTACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 236
QY 2990 CCTGAGGAGATATCCATATCCAGCCCTGTACTCCCGCAGCCTCATGTC 3049
DB 237 CCTTCT 296
QY 3050 AAGTCCGCTCTCTCTCC 3067
DB 297 CAGTGGCGTATCTCGNC 314

RESULT 13
PCT-US95-10202-24
Sequence 24, Application PC/TUS9510202

GENERAL INFORMATION:

APPLICANT: Shattuck-Eidens, Donna M.

APPLICANT: Sliard, Jacques

APPLICANT: Emi, Mitsuru

APPLICANT: Nakamura, Yusuke

APPLICANT: Durocher, Franck

TITLE OF INVENTION: In Vivo Mutations and Polymorphisms

TITLE OF INVENTION: In the 17q-Linked Breast and Ovarian Cancer

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP

STREET: 1201 New York Avenue, N.W., Suite 1000

NUMBER OF SEQ ID NOS: 2276
SOFTWARE: PatentIn version 3.0
SEQ ID NO 521
LENGTH: 4040
TYPE: DNA
ORGANISM: Homo sapiens
US-09-954-456-521

Query Match 100.0%; Score 4039.6; DB 10; Length 4040;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4039; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCTTCCACCCCTAGTCCAGGATCTGATACGGGAAACCCACACAGACGCCGGA 60
DB 1 GTCTTCCACCCCTAGTCCAGGATCTGATACGGGAAACCCACACAGACGCCGGA 60
QY 61 GCCCCCCACCCGTCAGAGCAACAGAGCCCGCTCTTGCTGGCTGAGATACATTG 120
DB 61 GCCCCCCACCCGTCAGAGCAACAGAGCCCGCTCTTGCTGGCTGAGATACATTG 120
QY 121 CTCTCTTGGTGAATCAGCTGTCCCTCTTGCGGAAATGAAACAGAAATCTT 180
DB 121 CTCTCTTGGTGAATCAGCTGTCCCTCTTGCGGAAATGAAACAGAAATCTT 180
QY 181 CGTGTCTCTATCTGTGCGGATCACCATCTTGCTGGTGTGCTGCTGCTGCTG 240
DB 181 CGTGTCTCTATCTGTGCGGATCACCATCTTGCTGGTGTGCTGCTGCTGCTG 240
QY 241 CAGGGGTGAGATGGGGGTAAACCCAGCCAGCTTCCCATTTGCCCTGTATCTCCAG 300
DB 241 CAGGGGTGAGATGGGGGTAAACCCAGCCAGCTTCCCATTTGCCCTGTATCTCCAG 300
QY 301 TGCCCAACCTTGAGACACACCTGGCCAGACCACTTTTGCAACCTGAGCCAGAGA 360
DB 301 TGCCCAACCTTGAGACACACCTGGCCAGACCACTTTTGCAACCTGAGCCAGAGA 360
QY 361 GCTGACGGCTGTGATGCGCTTTTGACCCAGCGCTGGGCGAGGGTGGTATGAGC 420
DB 361 GCTGACGGCTGTGATGCGCTTTTGACCCAGCGCTGGGCGAGGGTGGTATGAGC 420
QY 421 CAGAGCCCGGCTCGGACAACTGTCTTCTCAGTGAAGTTCAGCTGCTCCCAAGC 480
DB 421 CAGAGCCCGGCTCGGACAACTGTCTTCTCAGTGAAGTTCAGCTGCTCCCAAGC 480
QY 481 TGACGCTGTGCTACTTGAGACAGGGGAGGCCCCCACTGCGGGAGGACTGGCAT 540
DB 481 TGACGCTGTGCTACTTGAGACAGGGGAGGCCCCCACTGCGGGAGGACTGGCAT 540
QY 541 CGTCTTCTTGGAGGCAACCCAGCCCAAGCTGATGAGCTGGTGGTGGGCCACTGCC 600
DB 541 CGTCTTCTTGGAGGCAACCCAGCCCAAGCTGATGAGCTGGTGGTGGGCCACTGCC 600
QY 601 TCACCCCTCTCAATGCGGGAGCTGATGTGAGCGTCTATGGAGCCCGCTGCTATCA 660
DB 601 TCACCCCTCTCAATGCGGGAGCTGATGTGAGCGTCTATGGAGCCCGCTGCTATCA 660
QY 661 CCGAGGCGCGGTGTGTTCCAAAGATACCTGAGATAGACAGATCTTCAACAGAGA 720
DB 661 CCGAGGCGCGGTGTGTTCCAAAGATACCTGAGATAGACAGATCTTCAACAGAGA 720
QY 721 GCTGCCCCAGGCTTCTGGGCTTCTCAGCACTGTTGCTTCAACAGAGCCGGA 780
DB 721 GCTGCCCCAGGCTTCTGGGCTTCTCAGCACTGTTGCTTCAACAGAGCCGGA 780
QY 781 CCTGTGATGATGACACAGGCTCCCGGTGTGATCAGAGGACCGGGCCACTGTTT 840
DB 781 CCTGTGATGATGACACAGGCTCCCGGTGTGATCAGAGGACCGGGCCACTGTTT 840
QY 841 TGCGCTCTACTACAACATCTCGGGGCTGGTCTTCTGACACAGCGTGGGCTTGGACT 900
DB 841 TGCGCTCTACTACAACATCTCGGGGCTGGTCTTCTGACACAGCGTGGGCTTGGACT 900
QY 901 GCTAGTGAACCAAGGCGCTTGACCTTGCCCGCTGAGACTATCCAGAAAGTGTCTATCA 960

DB 901 GCTAGTGAACCAAGGCGCTTGACCTTGCCCGCTGAGACTATCCAGAAAGTGTCTATCA 960
QY 961 AGGCGGCTACTAGACAGACCTTGCCCAAGCTGAGGCGCACTTGGAGCCGGCTGTGAA 1020
DB 961 AGGCGGCTACTAGACAGACCTTGCCCAAGCTGAGGCGCACTTGGAGCCGGCTGTGAA 1020
QY 1021 TGTGTCTATCCAGACAAATGACAGTGGTCTGTGCTGTAAGTCCCTGTGCC 1080
DB 1021 TGTGTCTATCCAGACAAATGACAGTGGTCTGTGCTGTAAGTCCCTGTGCC 1080
QY 1081 CCGGGGTACAGCTCCCGCTTACAGTCTATCCCAAGGCGCGCTTCACTGATGCAAGG 1140
DB 1081 CCGGGGTACAGCTCCCGCTTACAGTCTATCCCAAGGCGCGCTTCACTGATGCAAGG 1140
QY 1141 AAGTGAAGTGGCTCTCACTGTGAGCTTCTCTTGGCTCGGAGCAATTCAGTGGCC 1200
DB 1141 AAGTGAAGTGGCTCTCACTGTGAGCTTCTCTTGGCTCGGAGCAATTCAGTGGCC 1200
QY 1201 AAGATCTTGGAGCTTGGCTTCCAAAGAAAGTATGTTATGAGTAAAGCTCCAAAGA 1260
DB 1201 AAGATCTTGGAGCTTGGCTTCCAAAGAAAGTATGTTATGAGTAAAGCTCCAAAGA 1260
QY 1261 GGCCTTGGCATATGATGAGAAATTCAGAGCAAGTATGAGCAACCGCTATGATGATG 1320
DB 1261 GGCCTTGGCATATGATGAGAAATTCAGAGCAAGTATGAGCAACCGCTATGATGATG 1320
QY 1321 AGGCTTGGCATGAGGCAAGTACACACGCGCTGACCGCTGGGGTGGAGTGGCCCTACTT 1380
DB 1321 AGGCTTGGCATGAGGCAAGTACACACGCGCTGACCGCTGGGGTGGAGTGGCCCTACTT 1380
QY 1381 GGCACCTAGGTGAGTGGGCACTCTTTGGAGTCCAGGCGCCCAAGAAATATAGTGA 1440
DB 1381 GGCACCTAGGTGAGTGGGCACTCTTTGGAGTCCAGGCGCCCAAGAAATATAGTGA 1440
QY 1441 TGCTTTTGTGTGTTGAACAGAAACAGGCGCTCCCGCTGGGCGACCACTAGATCT 1500
DB 1441 TGCTTTTGTGTGTTGAACAGAAACAGGCGCTCCCGCTGGGCGACCACTAGATCT 1500
QY 1501 CTACTGCGACTATTTGGGGGTCTTGGGAAACGGTCTGGTGTGATGATGATGATG 1560
DB 1501 CTACTGCGACTATTTGGGGGTCTTGGGAAACGGTCTGGTGTGATGATGATGATG 1560
QY 1561 CTTGCTCAACTATGATATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
DB 1561 CTTGCTCAACTATGATATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
QY 1621 ACGATCTATGCAAGGCTATCATCACTGCGCATCTCTTTGGTCTACTGGAAGTA 1680
DB 1621 ACGATCTATGCAAGGCTATCATCACTGCGCATCTCTTTGGTCTACTGGAAGTA 1680
QY 1681 CCGGAAACCAATGTCAAGCAACACCTGGGACGCTGACACCCAGCGCCCACTTCAA 1740
DB 1681 CCGGAAACCAATGTCAAGCAACACCTGGGACGCTGACACCCAGCGCCCACTTCAA 1740
QY 1741 GGTGATCTGATATGACAGAGCTGAGAACTGGGTGGGCGCAGAGATATGATGATG 1800
DB 1741 GGTGATCTGATATGACAGAGCTGAGAACTGGGTGGGCGCAGAGATATGATGATG 1800
QY 1801 CCCCATGGCTGTGCGCTGAGACCTTGAGACACACACTGACAGGCTGAGAGGAGCCGGA 1860
DB 1801 CCCCATGGCTGTGCGCTGAGACCTTGAGACACACACTGACAGGCTGAGAGGAGCCGGA 1860
QY 1861 GCTGCTGAGATGAGAGAGGAGGCGGCTTCTGCTGGGAAAGCCCAACCTCTCTACT 1920
DB 1861 GCTGCTGAGATGAGAGAGGAGGCGGCTTCTGCTGGGAAAGCCCAACCTCTCTACT 1920
QY 1921 GTACCTGGCGAGCAACCAAGTGGGCTGACCCCGGGGCTTACCGCATCAGAT 1980
DB 1921 GTACCTGGCGAGCAACCAAGTGGGCTGACCCCGGGGCTTACCGCATCAGAT 1980
QY 1981 GCTCAGCTTGTGAGAGCGGCTGCCCCAAGACGCTCCATGCGGAGAGGCTTCACTG 2040

Dp	1981	GCACACTTGTGCTGGAGAGCCGCTGGCCCCAAACAGCTCATGCGGAGGCTTCAGCTG	2040
Qy	2041	GGAGAGTACACAGCTGGCTGTGACCCAGGGAAGAGAGAGCCCACTACGACAGCT	2100
Dp	2041	GGAGAGTACACAGCTGGCTGTGACCCAGGGAAGAGAGAGCCCACTACGACAGCT	2100
Qy	2101	TTTCAATCAGAAATGACCTTGGGGCCCCACGTGATTTCAATGACTTCATCAACATGA	2160
Dp	2101	TTTCAATCAGAAATGACCTTGGGGCCCCACGTGATTTCAATGACTTCATCAACATGA	2160
Qy	2161	GACCATTTGCTGGAAGGATTTGGTGGCTTGGGTGACAGCTGGTTTTCGATATCCACA	2220
Dp	2161	GACCATTTGCTGGAAGGATTTGGTGGCTTGGGTGACAGCTGGTTTTCGATATCCACA	2220
Qy	2221	TGCAGAGACATTCCTAACACANAGTACTGTGGGGAAGGCGGGGCTTCCTCCGACG	2280
Dp	2221	TGCAGAGACATTCCTAACACANAGTACTGTGGGGAAGGCGGGGCTTCCTCCGACG	2280
Qy	2281	CTATTAATTTTGTGACGAAGACCCCTCTTCTACTGTGCGCACTCCATCTACTCCGAGG	2340
Dp	2281	CTATTAATTTTGTGACGAAGACCCCTCTTCTACTGTGCGCACTCCATCTACTCCGAGG	2340
Qy	2341	GGACAGAGATGTGGGGCCCGGAGGATCAACCCCTAGCTTCCTCGCCCAAGGCGCTGC	2400
Dp	2341	GGACAGAGATGTGGGGCCCGGAGGATCAACCCCTAGCTTCCTCGCCCAAGGCGCTGC	2400
Qy	2401	CTGTGCCCCGACCTCCCTGCTTCCTCCAGGCGGGCTTCCTCAACTAGGCGGCTCT	2460
Dp	2401	CTGTGCCCCGACCTCCCTGCTTCCTCCAGGCGGGCTTCCTCAACTAGGCGGCTCT	2460
Qy	2461	GGATGGGGGATGTGGGCAAGGGGCTCCAGGGCCAGGGGTATGAGGATGGGAGACGTGG	2520
Dp	2461	GGATGGGGGATGTGGGCAAGGGGCTCCAGGGCCAGGGGTATGAGGATGGGAGACGTGG	2520
Qy	2521	GCATGGGGCGGAGGCTGGTTCCTCTTCTGTCGACAGACTCTTTTCTTCACACTAC	2580
Dp	2521	GCATGGGGCGGAGGCTGGTTCCTCTTCTGTCGACAGACTCTTTTCTTCACACTAC	2580
Qy	2581	CCTCCCTGCGATCGGCTCTGAGCCAGGAGGCTCTGAGCCCTGATGCGCTGACACAGG	2640
Dp	2581	CCTCCCTGCGATCGGCTCTGAGCCAGGAGGCTCTGAGCCCTGATGCGCTGACACAGG	2640
Qy	2641	GACACTAAACCTTGTGATCCGACGCTGACTGATGATTTCTATCCACAGAGGCCAGGACTGG	2700
Dp	2641	GACACTAAACCTTGTGATCCGACGCTGACTGATGATTTCTATCCACAGAGGCCAGGACTGG	2700
Qy	2701	CCGAGCTGGAGCGGTGGCGGAGGAGGCTTCCCTAGATGATTCCTTTGTGTGCTGTGGCT	2760
Dp	2701	CCGAGCTGGAGCGGTGGCGGAGGAGGCTTCCCTAGATGATTCCTTTGTGTGCTGTGGCT	2760
Qy	2761	TTCCCGAATCTTTTATAGCCCACTCCCAAGGACTCTTAAAGGGGGCTATTTCCCTGAGAC	2820
Dp	2761	TTCCCGAATCTTTTATAGCCCACTCCCAAGGACTCTTAAAGGGGGCTATTTCCCTGAGAC	2820
Qy	2821	CCAGAGTAGGGTTGCGACGTCTGCAAGTCCATAGCTGAGACTGGAAAGAGATGCTTGTGCTC	2880
Dp	2821	CCAGAGTAGGGTTGCGACGTCTGCAAGTCCATAGCTGAGACTGGAAAGAGATGCTTGTGCTC	2880
Qy	2881	ACATTCCTCTCATCCAGGTCTTTCCTTCTGCTTCTCTCTCTCACTACTTCTTCC	2940
Dp	2881	ACATTCCTCTCATCCAGGTCTTTCCTTCTGCTTCTCTCTCTCACTACTTCTTCC	2940
Qy	2941	TCGCTCCCGCTTCTGCGGCTTCTCTTCAACCGCAATTCGCCGAATCCGAGGGGAT	3000
Dp	2941	TCGCTCCCGCTTCTGCGGCTTCTCTTCAACCGCAATTCGCCGAATCCGAGGGGAT	3000
Qy	3001	ATCCCATATGTCAGGCCCTGTGATCTCCGCCAGCCCTCAGTTTTCAGTCAATGCTCCGCT	3060
Dp	3001	ATCCCATATGTCAGGCCCTGTGATCTCCGCCAGCCCTCAGTTTTCAGTCAATGCTCCGCT	3060
Qy	3061	CCCTTCAGGCCCTATGGAATCTTCAAGTCAAGGAGCCCTTATCAGAGTGGCCAAATCC	3120
Dp	3061	CCCTTCAGGCCCTATGGAATCTTCAAGTCAAGGAGCCCTTATCAGAGTGGCCAAATCC	3120

QY	3121	TGATGACGCTCCCTGTGTGCTGCTTACTTAATGGAGATGAGAGTGGCTACACCCCTGC	3180
QY	3121	TGATGACGCTCCCTGTGTGCTGCTTACTTAATGGAGATGAGAGTGGCTACACCCCTGC	3180
QY	3181	CTGGGGCTGGGCTGTGTTTCAGGACAGCTCTCTGTGCAATTTGTCGTCCGTGCATG	3240
Db	3121	TGATGTCGTGCTCCCTGTGTGCTGCTTACTTAATGGAGATGAGAGTGGCTACACCCCTGC	3180
QY	3181	CTGGGGCTGGGCTGTGTTTCAGGACAGCTCTCTGTGCAATTTGTCGTCCGTGCATG	3240
Db	3181	CTGGGGCTGGGCTGTGTTTCAGGACAGCTCTCTGTGCAATTTGTCGTCCGTGCATG	3240
QY	3241	CTCTCTATAGAGAGAGATGGTCAATGCTGACACAGCAGCTCAAGTTACATTTCAAGTGA	3300
Db	3241	CTCTCTATAGAGAGAGATGGTCAATGCTGACACAGCAGCTCAAGTTACATTTCAAGTGA	3300
QY	3301	TTTGGGGGCTGCATGATTAATGAAGATGGCCATTTTGTACCGAGGGCTGTGATTTCTGAA	3360
Db	3301	TTTGGGGGCTGCATGATTAATGAAGATGGCCATTTTGTGACCGAGGGCTGTGATTTCTGAA	3360
QY	3361	CAGCCTTTTGGAGGCTGAGATGTGAACAAAGGGTGGGCATCAAAAGATGAGAACCCAA	3420
Db	3361	CAGCCTTTTGGAGGCTGAGATGTGAACAAAGGGTGGGCATCAAAAGATGAGAACCCAA	3420
QY	3421	GCCCTTACAACTCCAGCCACCCAGCAGAGGGGGCTGTCCATTCACATTTACAGCATGGGA	3480
Db	3421	GCCCTTACAACTCCAGCCACCCAGCAGAGGGGGCTGTCCATTCACATTTACAGCATGGGA	3480
QY	3481	ATGAGCTGGGGCCCGGGGTGAGTGGGGGTCGTGGCCATGAGGGAGGGGCTGGGCTGGGT	3540
Db	3481	ATGAGCTGGGGCCCGGGGTGAGTGGGGGTCGTGGCCATGAGGGAGGGGCTGGGCTGGGT	3540
QY	3541	GGGGCAGGGGCTGGGCTGGTCCAGGCTTGGGCTCCATTTCCATCCTGCTGCTCCCTG	3600
Db	3541	GGGGCAGGGGCTGGGCTGGTCCAGGCTTGGGCTCCATTTCCATCCTGCTGCTCCCTG	3600
QY	3601	AGGCTGTGATTTGGGATGGGGACAAAGAAATAGCAAGAGATGAGAAACAAGAACTTT	3660
Db	3601	AGGCTGTGATTTGGGATGGGGACAAAGAAATAGCAAGAGATGAGAAACAAGAACTTT	3660
QY	3661	TTTCTCAAAGGACTGGTTAAATCAATTCGTGATACAGCTTACATTAATATGATGAG	3720
Db	3661	TTTCTCAAAGGACTGGTTAAATCAATTCGTGATACAGCTTACATTAATATGATGAG	3720
QY	3721	CTAAAAATATATGTATGCTTTTATATACTAATATGTAATATCTTCAGGTGAAAAAGC	3780
Db	3721	CTAAAAATATATGTATGCTTTTATATACTAATATGTAATATCTTCAGGTGAAAAAGC	3780
QY	3781	AAGCCACAGAAATGTATATAGCCGACTTCCCATTTGTGTTACAGAAAGAGTAGAATATA	3840
Db	3781	AAGCCACAGAAATGTATATAGCCGACTTCCCATTTGTGTTACAGAAAGAGTAGAATATA	3840
QY	3841	AACACATATATTCCTATAGTATGGCTAATTCAGATTAATGGGTAACACATGATTACTTTGG	3900
Db	3841	AACACATATATTCCTATAGTATGGCTAATTCAGATTAATGGGTAACACATGATTACTTTGG	3900
QY	3901	GAGGGGAACAGTAGTGTGAGGACAGAGAGGGAAGGCTTAACACTTACACCCCTTTG	3960
Db	3901	GAGGGGAACAGTAGTGTGAGGACAGAGAGGGAAGGCTTAACACTTACACCCCTTTG	3960
QY	3961	TACATTTTGAATTTTGAACCATGTGACTGATTACTTAATTAATTAACATTAATGGG	4020
Db	3961	TACATTTTGAATTTTGAACCATGTGACTGATTACTTAATTAATTAACATTAATGGG	4020
QY	4021	CCCAAAAAAAAAAAAAAAAA 4040	
Db	4021	CCCAAAAAAAAAAAAAAAAA 4040	

```

RESULT 3
US-10-081-408-19
; Sequence 19, Application US/10081408
; Patent No. US20020160482A1
; GENERAL INFORMATION:
; APPLICANT: Abrams, n, Lars
; APPLICANT: Nilsson, Joakim

```


;; TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
;; FILE REFERENCE: 13425-03501
;; CURRENT APPLICATION NUMBER: US/10/081,408
;; CURRENT FILING DATE: 2002-02-21
;; PRIOR APPLICATION NUMBER: SE 0100625-3
;; PRIOR FILING DATE: 2001-02-23
;; PRIOR APPLICATION NUMBER: US 60/272,247
;; PRIOR FILING DATE: 2001-02-28
;; NUMBER OF SEQ ID NOS: 20
;; SOFTWARE: FASTSEQ for Windows Version 4.0
;; SEQ ID NO 19
;; LENGTH: 3006
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Recombinant construct
;; NAME/KEY: CDS
;; LOCATION: (1)...(2994)
US-10-081-408-19

Query Match 54.7% Score 2209.2; DB 9; Length 3006;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2211; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 245 GGTGAGATGGGGTGAACCCAGCAGCTTCCCATTTGCCCTCTGTATCTCCAGTGC 304
DB 790 GGTGAGATGGGGTGAACCCAGCAGCTTCCCATTTGCCCTCTGTATCTCCAGTGC 849
QY 305 CAGCTTGGACACACCTGGCCAGAGCAGCTGTTTGACAGCTAGAGCCGAGAGAGCTG 364
DB 850 CAGCTTGGACACACCTGGCCAGAGCAGCTGTTTGACAGCTAGAGCCGAGAGAGCTG 909
QY 365 ACGGCTGTGATGCGCTTTCTGACCCAGGCGTGGGGCCAGGGGCTGTGATGACCCAG 424
DB 910 ACGGCTGTGATGCGCTTTCTGACCCAGGCGTGGGGCCAGGGGCTGTGATGACCCAG 969
QY 425 GCCCGGCCCTGGACAACTGTCTTCTCACTGAGTTGACAGTCCCTCCCAAGCTGCA 484
DB 970 GCCCGGCCCTGGACAACTGTCTTCTCACTGAGTTGACAGTCCCTCCCAAGCTGCA 1029
QY 485 GCCCGGCCCTGGACAACTGTCTTCTCACTGAGTTGACAGTCCCTCCCAAGCTGTC 544
DB 1030 GCCCGGCCCTGGACAACTGTCTTCTCACTGAGTTGACAGTCCCTCCCAAGCTGTC 1089
QY 545 TTCTTGGCAGGCAACCCAGCAGCTGAGTGAAGTGTGAGTGGGCTGGGCTGCTGCTAC 604
DB 1090 TTCTTGGCAGGCAACCCAGCAGCTGAGTGAAGTGTGAGTGGGCTGGGCTGCTGCTAC 1149
QY 605 CCTCTCACTGAGGAGCTGAGTGTGAGGCTCATGAGGCCCCCTGCCCTATCACCGA 664
DB 1150 CCTCTCACTGAGGAGCTGAGTGTGAGGCTCATGAGGCCCCCTGCCCTATCACCGA 1209
QY 665 CGCCCGGCTGTTCACAGAGTACCTGAGCATAGACAGATGATTTCAACAGAGCTG 724
DB 1210 CGCCCGGCTGTTCACAGAGTACCTGAGCATAGACAGATGATTTCAACAGAGCTG 1269
QY 725 CCCCAGGCTTCTGGGCTTCTGCACACACTGTGCTTCTACAAAGCACCGGGAGCGAACTG 784
DB 1270 CCCCAGGCTTCTGGGCTTCTGCACACACTGTGCTTCTACAAAGCACCGGGAGCGAACTG 1329
QY 785 GTGACAATGACACAGGCTCCCGGTGCTGCAATCAGGGGACCGGGCCACCGGTTTGGC 844
DB 1330 GTGACAATGACACAGGCTCCCGGTGCTGCAATCAGGGGACCGGGCCACCGGTTTGGC 1389
QY 845 CTCTACTACAACTCTCGGGGCTGGGCTTCTTCTGACACACAGTGGGCTGGAGCTGTA 904
DB 1390 CTCTACTACAACTCTCGGGGCTGGGCTTCTTCTGACACACAGTGGGCTTGGAGCTGTA 1449
QY 905 GTGAACCAAGGCCCTTGAACCTGCGCGCTGAGACTATCAGAGAGGTGTTATCAAGGC 964
DB 1450 GTGAACCAAGGCCCTTGAACCTGCGCGCTGAGACTATCAGAGAGGTGTTATCAAGGC 1509
QY 965 CGCTACTAGCAGAGCTGGCCAGCTGAGAGGCCAGTTTGAAGCGCGGCTGTGATGTG 1024

DB 1510 CGCTACTAGCAGAGCTGGCCAGCTGAGAGCCAGAGTTTGAAGCGCGGCTGTGATGTG 1569
QY 1025 GTGCTGATCCAGACAATGAGCAGTGGGCTGTGCTGCCGGAAGTCCCTGTGCCCGC 1084
DB 1570 GTGCTGATCCAGACAATGAGCAGTGGGCTGTGCTGCCGGAAGTCCCTGTGCCCGC 1629
QY 1085 GGTCCAGCTCCCTCTACAGTCTATCCCAAGGCCCGCTTCACTGATCCAGGAGT 1144
DB 1630 GGTCCAGCTCCCTCTACAGTCTATCCCAAGGCCCGCTTCACTGATCCAGGAGT 1689
QY 1145 CGATGGGCTCTCACTGTGAGTCTTCTCTTGGGCTCGGAGCATTCAGTGGCCCAAG 1204
DB 1690 CGATGGGCTCTCACTGTGAGTCTTCTCTTGGGCTCGGAGCATTCAGTGGCCCAAG 1749
QY 1205 ATCTTGAAGTCTTCCCTTCCAGAGAGAAAGACTAGTTATGATTAAGCTCCAGAGGC 1264
DB 1750 ATCTTGAAGTCTTCCCTTCCAGAGAGAAAGACTAGTTATGATTAAGCTCCAGAGGC 1809
QY 1265 TTGGCATCTATGTTGGAATTTCCCAAGCAGCAATGACGACCGCTATGTTGATGAGGC 1324
DB 1810 TTGGCATCTATGTTGGAATTTCCCAAGCAGCAATGACGACCGCTATGTTGATGAGGC 1869
QY 1325 TTTGGCATGGGCAATGACACGCGCTGACCGCTGGGCTGAGTGGCTCTACTTGGCC 1384
DB 1870 TTTGGCATGGGCAATGACACGCGCTGACCGCTGGGCTGAGTGGCTCTACTTGGCC 1929
QY 1385 ACCTACGAGTGGACACTTCTTTTGGAGTCCAGAGCGCCCAAGACAAATACGATGCC 1444
DB 1930 ACCTACGAGTGGACACTTCTTTTGGAGTCCAGAGCGCCCAAGACAAATACGATGCC 1989
QY 1445 TTTTGTGTGTTGAACAGAACAGAGGCTCCCTCGCGGACACACTGATCTGTAC 1504
DB 1990 TTTTGTGTGTTGAACAGAACAGAGGCTCCCTCGCGGACACACTGATCTGTAC 2049
QY 1505 TCGCATCTATTGGGGGCTTGGGAAACGGTGCAGTGCAGATCTATCCACCTG 1564
DB 2050 TCGCATCTATTGGGGGCTTGGGAAACGGTGCAGTGCAGATCTATCCACCTG 2109
QY 1565 CTCAACTATGACTATGTTGGGATACGGTCTTCCACCCCAAGGGCCATAGAAATACGA 1624
DB 2110 CTCAACTATGACTATGTTGGGATACGGTCTTCCACCCCAAGGGCCATAGAAATACGA 2169
QY 1625 TTCTATGCGCAGGGCTATATAGCTCGGCAATTCCTTGTGCTACAGGGAATACGGG 1684
DB 2170 TTCTATGCGCAGGGCTATATAGCTCGGCAATTCCTTGTGCTACAGGGAATACGGG 2229
QY 1685 AACCAAGTGTAGAGACACACCTGGGACAGGCTCCACACCCAGAGGCCCACTTAAGTG 1744
DB 2230 AACCAAGTGTAGAGACACACCTGGGACAGGCTCCACACCCAGAGGCCCACTTAAGTG 2289
QY 1745 GATCTGATGTAGCAGAGCTGGAAGTGGGTCTGGGCCGAGATATGCTTTGTCCCC 1804
DB 2290 GATCTGATGTAGCAGAGCTGGAAGTGGGTCTGGGCCGAGATATGCTTTGTCCCC 2349
QY 1805 ATGGCTGTGCCCTGGAGCCCTGAGCAGCAGTGCAGAGAGCTGACAGTCAACCGGAAGTG 1864
DB 2350 ATGGCTGTGCCCTGGAGCCCTGAGCAGCAGTGCAGAGAGCTGACAGTCAACCGGAAGTG 2409
QY 1865 CTGAGATGAGAGGAGCAGGCGGCTCTCTGTTGGGAAACGCGACCGCTCGCTACTGTAC 1924
DB 2410 CTGAGATGAGAGGAGCAGGCGGCTCTCTGTTGGGAAACGCGACCGCTCGCTACTGTAC 2469
QY 1925 CTGGCAGACCAACAGCAGCAAGTGGGCTACACCCCGGGCTACCGCATCAGATGCTC 1984
DB 2470 CTGGCAGACCAACAGCAGCAAGTGGGCTACACCCCGGGCTACCGCATCAGATGCTC 2529
QY 1985 AGCTTGTCTGAGAGCCGCTGCCAAAACAGCTCATGCGAGAGGCTTACGCTGGAG 2044
DB 2530 AGCTTGTCTGAGAGCCGCTGCCAAAACAGCTCATGCGAGAGGCTTACGCTGGAG 2589
QY 2045 AGGTACAGAGTGGCTGTACCCAGGAGGAGGAGGAGGCCAGTACAGCAGGTTTC 2104

Db	Sequence	Score	DB ID	Length	Indels	Gaps
Db	ATGAGCTGCTGGAGCGTCATGCGGCCCCCTGCCCCCTATTCACGGCGTCTGCTGACCA	605				
Qy	681 AAGAGTACCTGTGACATAGACACCATGATCTTCAACAGAGAGCTGCCACGGCTTGTGGGC	740				
Db	GAGAGTATGACGATATTCAGAGATGATCTTTCACAGAGAGCTGCCACACGGCTGTGTC	665				
Qy	741 TTCTCCACCACTGTTGCTTTCACAAAGCAGCGGGAGCGGAACCTGGTGACATGACCAAG	800				
Db	TCTTCATCTACTGTGCTTCTTACAAACGCCAAGGACACCTGCTAATAAATGACTACAG	725				
Qy	801 CTCCCCGATGCTTGTCAATCAGAGGACCGGGCCACCTGGTTTGGCCCTTACTACAAACATCT	860				
Db	726 CCCCCCGGTGTTGCAATCAGAGGACCGGGCCACCTGGTTTGGCATATATTACATATCT	785				
Qy	861 CGGGCGCTGGGTTCTCTCTGACCAACCTGGGCTTGGAGACTCTAGTGACCAACAAAGGCC	920				
Db	786 CAGGGGCTGGGTTTATCCCTACCCCATTTGGCTTGAAGCTTCTGTATATACAAAGGCC	845				
Qy	921 TTGACCCCTGGCCCGCTGAGACTATCCAGAGGTTCTTATCAAGAGCCGCTACTACGACACC	980				
Db	846 TGAATCTCGCCCTGTGAGACATCCAGAGGATTTCTTACAAAGGCCGCTACTATGAGATC	905				
Qy	961 TGGCCCACTGTGAGGCCCCAGTTTGAAGCCCGCCCTGTGTGATGTGCTGATATCCACACA	1040				
Db	906 TGACTCACCTGTGAGGACATGTTTGAAGCTGGCTGTGTAATGTGTTTGGTCCGACACA	965				
Qy	1041 ATGGACAGGTGGGTCCTGAGTCCCTGAGAGTCCCTGTGTCGCCCGGGTGCATCCGCCCTC	1100				
Db	966 ATGGACAGGTGGGTCCTGAGTCCCTGAGAGTCCCTGTGTCGCCCGGGTGCATCCGCCCTC	1025				
Qy	1101 TACAGTCTATCCCAAGGCCCGC 1125					
Db	1026 TGCARTTCAYCACNGARGNCNMG 1050					

RESULT 6
 US-09-962-832-126
 / Sequence 126, Application US/09962832
 / Patent No. US20020110821A1
 / GENERAL INFORMATION:
 / APPLICANT: Ebner, Reinhard
 / TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signature
 / TITLE OF INVENTION: Sets
 / FILE REFERENCE: 688290-74
 / CURRENT APPLICATION NUMBER: US/09/962, 832
 / PRIOR FILING DATE: 2001-09-25
 / PRIOR APPLICATION NUMBER: US/60/235,077
 / PRIOR FILING DATE: 2000-09-25
 / PRIOR APPLICATION NUMBER: US/60/235,280
 / PRIOR FILING DATE: 2000-09-25
 / NUMBER OF SEQ ID NOS: 259
 / SOFTWARE: PatentIn version 3.0
 / SEQ ID NO 126
 / LENGTH: 2441
 / TYPE: DNA
 / ORGANISM: Homo sapiens
 US-09-962-832-126

Query Match 10.5%; Score 422.2; DB 10; Length 2441;
 Best Local Similarity 52.8%; Pred. No. 2.9e-115;
 Matches 1099; Conservative 0; Mismatches 918; Indels 66; Gaps 6;

Db	Sequence	Score	DB ID	Length	Indels	Gaps
Db	CCACGACCAATGGCCAGAAACACCGTGTCTTCATGAGATGCTGTGCTCCCAAGATACC	339				
Qy	423 AGGCCGGGCGCTGGACAACATGTGTCTTCAGTGAAGTTGCACTGCTCCCAAGGCTG	482				
Db	220 TGAAGCAATGACAGACCTTCTCTGTGTCACAAAGAGAGTGAAGCTGACGCCCTCCAGTA	279				
Qy	363 TGACGGCTTGAATGACCTTTTGACCAACGCGCTGGGCCAGGCTGTGTGATGACGCC	422				
Db	160 CCTCCCGGGGACTGCCCCAGGAAGGACGAGGGTGTTCACACTTAACCAACAGAGAC	219				
Qy	303 CCCAGCTTGGACACACCTGTGCCAGAGCAGCAGCTGTTTGCACAACTGAGCCGAGAGAC	362				
Db	160 CCTCCCGGGGACTGCCCCAGGAAGGACGAGGGTGTTCACACTTAACCAACAGAGAC	219				

Oy	483	CAGCCCTGCGTCACTTGGACAGGGGAGACCCCCACACTGTCGCCGGAGGCACTGGCCATCG	542
Db	340	ATGTCAGAGGTTCTTGGATAAAGGTAAGGCACTCTGTGGGAGAACCCGTGCTCGTCA	399
Oy	543	TCTCTTTGGCAGGCAACCCAGCCCAACGTAGTAGTGGTGTGGGGCACTGGCTC	602
Db	400	TCTTCTTTGGTGAACAGAGACATCCCAATGTCAAGAGTTTGTGTGGGGCCCTGGCAG	459
Oy	603	ACCCCTCTACATGCGGGAGCTGACTGTGAGC---GTCAATGAGGCCCCCTGCCATTC	659
Db	460	GGCCCTCTACATGCGAGCACTGTCCCGAGGCTGGGTACCAAGTCTCTCGGCAATCGA	519
Oy	660	ACCCAGCCCCGGTGTCTTCCAAAGATACCTGGACATACAGATATGATCTTCAACAGAG	719
Db	520	GGCCCATCTCCACAGCAGAGATATGCCCTCTCTACACACACCTCGAGGAACCAACAGC	579
Oy	720	AGCTGCCCAAGGCTTCTGGGCTTCTCCACACTGTGTCTTACAAAGCAGCGGGAGCGA	779
Db	580	CCCTGATCATGTTCTTCTCCATACCAAGGCTTCTCATTCMAACTGCCATGACAGAT	639
Oy	780	ACCTGTGACAAATAGCACAGGCTCCCCGTGTCTGTGAATACAGGGAACGGGGCACTGGT	839
Db	640	GCTGTGGCTTACCGATGTGGCCCCCGGGGTGTGGCTTGTGCAAGCGCCGCAATGTGC	699
Oy	840	TTGGCCTCTACTACACATGTGCGGGCGTGGGTTCTTCTCGACACAGCTGGGCTGGAGC	899
Db	700	TTATCATACAGCGGTATGTGAAGGC-----TACTTTGTGACCCCACTGGGCTGGAGC	753
Oy	900	TGCTAGTGAACACACAGGCGCTTGACACCTTGCCCGCTGGACTATCCAAAGTGTTCATC	959
Db	754	TCTCTGTGGATCATGGGAGACAGATGTGGGCACTGGCCGTGGAGCAGGTGTGTCA	813
Oy	960	AAGCGCGTATATACAGAGCGTGGCCAGCTGGAGGCCAGTTGTGGCGCGGCTGGTGA	1019
Db	814	ACGGGAAGTTTATGGAAGCCGACAGAGAACTGGCTGGAAATATGCATGTGAGAGGTGG	873
Oy	1020	ATGTGTGCTGATCCAGACA-----ATGGCAGAGTGGGT	1055
Db	874	ACGTGTGTCTCTGAGAGACCCGCTGCTGGGGCAAGGGGCATGACAGCACAGAGAGC	933
Oy	1056	CCTGTCCCTGAATCCCGTCGTGGCCCGGGGTCCACTCC-----	1095
Db	934	CGCCCTTTTCTCTCCACAGAGCCCGGGGACTTCCCAAGCCCATCATGTGAGCG	993
Oy	1096	-CCCTCTACAGTTTATCCCAAGGCCCCGCTTCAGTGTCCAGGGAAGTCCAGTGGCT	1154
Db	994	GGCCCCGCTGTGTCACAGCCCAAGGGCCGTGCTTCAAGGTGAGGGCAACGCTGTCT	1053
Oy	1155	CCTACTGTGCACTTCTCCTTTTGCCCTCGGAGCATTCAGTGGCCCAAGATCTTTGAC	1214
Db	1054	ACGGGCGGTGAGGTTTGGCTTCCGGGTGGGCTCTCTTCGGGGTCGAGGCTCTGAAGC	1113
Oy	1215	TTCCGTTTCCAGGAAAGACTATATTAGATTAAGACCTCCAAAGAGGCTTGGGCATCT	1274
Db	1114	TGCATTTGGGGGAGAGCGCATTTGCCATATAGGTCAAGCGTGAAGAGCAAGTGGCGCTGT	1173
Oy	1275	ATGTGTGAATTTCCCAAGCAGCAATGACGACCCGCTATGTGTGATGAGGCTTTGGCATGG	1334
Db	1174	ATGTGAGGACACACACTGTGAGGCAATGACAGCAAGTACCTCGATGTGCGGCTGGGGCTGG	1233
Oy	1335	GCATATACACAGCCCCCTGACCCCGTGGGTGAGACTGCCCCATCTTGGGCACATCGTGG	1394
Db	1234	GCAGGCTACATCATGAGTTAGCCCCCGGCAATCACTGCCCCGGAGACCGCACATTCCTGG	1293
Oy	1395	ACTGTGCACTTCTTTTGGAGTCCAGGCGCCCAAGAACATATCGTATGCTTTGTGT	1454
Db	1294	ACACTTTCACATACATATGATGTGCGATGACCCGGTCCATTATCCCGAGCGCTGTGCTCT	1353
Oy	1455	TTGAACAGACACAGGCGCTCCCGCTGGCGGCAAC-----ACTCAGATCTTACT	1505
Db	1354	TTGAATATCCCAACAGGGGTGGCCCTTGTGGGGGCACTTAAATTCACATTTAAAGTGTCT	1413

```
QY 1506 CGCACTACTTTGGGGGCTCTTGGGAAACGGTGTGTCAGATATATGTCACCTTGC 1565
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1414 TCACTTCTATTCGGGGCTGAAAGGGCCAGTCTGTGTGGGACACTTAACTGTCT 1473
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1566 TCACTATGACTATGTGTGGGATGAGTCTTCCACCCAGTGGGGCCATAGAAATACGAT 1625
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1474 ACAATTATGATATGTTGGGACTTATCTTACCCCAACGGGGGATGATGAGGCCAAGA 1533
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1626 TCTATGCCACGGGCTACATCACTGCGCATTCCTTTGTGCTCTACTGGGAAGTACGGGA 1685
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1534 TGCATGCCACATGGGCTACCTCCACGGCCACTTCTACACCCCGAGGGGGCTGCCACGGCA 1593
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1686 ACCAAGTGTCAAGACACACCTGGGACGGTCCACACCCACAGCCGCCACTTCAAGGTGG 1745
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1594 CTCGCCCTCACACACCACTGATTTGGCAACATACACTTACTTGTGCACTACCGCGTAG 1653
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1746 ATCTGATGTAGACAGACTGGAATGAGTCTGGCCGAGATATGCTTTGTGCCCA 1805
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1654 ACCTGATGTGGACGACCAAGAACAGCTTCCAGACACTGCAATGAAGTAAAGTAAAGAA 1713
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1806 TGGCTGTGCTGGAGCCCTGAGACACAGCTCAGAGGCTGCAAGTACCCGGAAGCTGC 1865
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1714 TCACCAACCCCTGGAGCCCAAGACACGGGTGTCTCAGCAACTGTGGACAGACGCACT 1773
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1866 TGGAGATGGAGAGCAGCCGCTCTCTGTGGAAAGGCGCACCCCTGCTACCTGTACC 1925
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1774 ACTCCTGGAGGGCGGAGCGGCGCTTCCGCTTAAAGAAAGTGCCTTAAGTACCTGTCT 1833
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1936 TGGCAGCAACCAAGCAAGAGTGGGTCAACCCCGGGGCTACCGCATCCAGATGCTCA 1985
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1834 TTACAGAGCCCGAGAGAGAACCCCTGGGGGACACAGCGAGCTACCGGCTGACAGTCACT 1893
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1986 GCTTGTGTGAGAGAGCCGCTGCCCAAAACAGTCCATGGCGAGAGCTTCACTGGGAGA 2045
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1894 CCATGGCCGACAGGATGTGCTGCCCAAGGCTGGCAGAGAGAGAGCCATCACTGGGCAA 1953
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2046 GSTACACAGTGTGACCCAGCAGGAAAGAGAGAGAGCCAGTACAGACAGCTTTTCA 2105
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1994 GSTACCCCTGGCAGTACAGCAAGTACCGGAGTCCGAGCTGTGCAGACAGACATCTACC 2013
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2106 ATCAGATAGACCTTGGGCCCCCACTGTGATTTCACTGACTTCAT--CAACAATAGA 2162
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2014 ACCAAGACAGACCCCTGGCAGCCGCGGTGTGTAGAGAGTCTTCCACAAACAGAGA 2073
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2163 CCATTGTGGAAGGATTTGGTGGCTGGTGACAGCTGTCTTGTGCAATCCACATG 2222
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2074 ACATTGAAAATGAGGACCTGGTGGCTGTGAGCGTGGGCTTCTGTCACATCCCCACT 2133
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2223 CAGAGACATTCCTTAACACAGTGTGTGGGAGAGCGGTGCTTCTTCCGACGCT 2282
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2134 CAGAGACATTCCTCAACACAGCAGACACTGGAACTCCCTGGGCTTCTGCTCGGCACT 2193
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2283 ATACTTCTTTGACGAGAACCCCTCTCTTCTACTGTGCCGACTC 2325
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2194 TCACTTCTTCCAGAGAGACCCCTCCCTGGCATCCAGAGACAC 2236
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
US-09-918-995-30382
: Sequence 30382, Application US/09918995
: Publication No. US20030073623A1
: GENERAL INFORMATION:
: APPLICANT: Hyseq, Inc.
: TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
: TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
: FILE REFERENCE: 20411-756
: CURRENT APPLICATION NUMBER: US/09/918, 995
: CURRENT FILING DATE: 2001-07-30
: PRIOR APPLICATION NUMBER: US/09/235, 076
: PRIOR FILING DATE: 1999-01-20
: NUMBER OF SEQ ID NOS: 38054
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 30382
```

```
: LENGTH: 459
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(459)
: OTHER INFORMATION: n = A,T,C or G
US-09-918-995-30382
```

```
Query Match 9.9%; Score 401.4; DB 9; Length 459;
Best Local Similarity 95.0%; Pred. No. 1.5e-109;
Matches 414; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
```

```
QY 821 GGGGACCGGCGCACTGTTTGGCTCTACTACAAATCTGGGGCTTCTCTCTG 880
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 23 GNGGAATTCGCCCACTAGTTTGGCTCTACTACAAATCTGGGGCTTCTCTCTG 82
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 881 CACCAAGTGGGCTTGGAGCTGCTAGTGAACCAAGGCGCTTACCCGCGGAGCT 940
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 83 CACCAAGTGGGCTTGGAGCTGCTAGTGAACCAAGGCGCTTACCCGCGGAGCT 142
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 941 ATCCAGAAAGTGTCTATCAAGGCGCTACTACAGACGCTGGCCAGTGGAGCCGAG 1000
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 143 ATCCAGAAAGTGTCTATCAAGGCGCTACTACAGACGCTGGCCAGTGGAGCCGAG 202
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1001 TTTGAGGCGGCGCTGTGTAATGTGCTGATCCAGACATGCGACAGTGGTCTTG 1060
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 203 TTTGAGGCGGCGCTGTGTAATGTGCTGATCCAGACAAAGGCAAGTGGTCTTG 262
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1061 TCCGTAAGTCCCTGTGCGCCCGGGGTCAGCTCCCGCTACATGTTATCCCAAGC 1120
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 263 TCCGTAAGTCCCTGTGCGCCCGGGGTCAGCTCCCGCTACATGTTATCCCAAGC 322
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1121 CCCGCTTCAGTGTCCAGGAAAGTGAAGTGGCTCTCACTGTGAGCTTCTCTTGGC 1180
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 323 CCCGCTTCAGTGTCCAGGAAAGTGAAGTGGCTCTCACTGTGAGCTTCTCTTGGC 382
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1181 CTCGAGCATTTAGTGGCCCAAGATCTTTCAGCTTCCCTTCCAGGAAAGATAGT 1240
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 383 CTCGAGCATTTAGTGGCCCAAGATCTTTCAGCTTCCCTTCCAGGAGAGGAGTGGCC 442
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1241 TATGAGATTAAGCTCC 1256
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 443 TATGAGATTAAGCTCC 458
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 8
US-09-960-352-6153
: Sequence 6153, Application US/09960352
: Patent No. US20020137139A1
: GENERAL INFORMATION:
: APPLICANT: Warren, Wesley C.
: APPLICANT: Tao, Nengping
: APPLICANT: Byatt, John C.
: APPLICANT: Mathialagan, Nagappan
: TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
: TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
: FILE REFERENCE: 16511.006/37-21(10298)C
: CURRENT APPLICATION NUMBER: US/09/960, 352
: CURRENT FILING DATE: 2001-09-24
: NUMBER OF SEQ ID NOS: 15112
: SEQ ID NO 6153
: LENGTH: 444
: TYPE: DNA
: ORGANISM: Bos taurus
: OTHER INFORMATION: Clone ID: 26-LIB34-078-Q1-E1-G9
US-09-960-352-6153

Query Match 7.8%; Score 315; DB 10; Length 444;
Best Local Similarity 82.8%; Pred. No. 1.2e-83;
Matches 360; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
QY 1293 CAGCAATGACGACCCGCTATGTGATGAGGCTTTGGCATGGGCAAGTACACAGCCGCC 1352
```


Db	136	AGTATGTTCTTCAAGCTGTGTGCGAGATGCTCAAAATATGACTTAATGTGGGATATATGTT	195
QY	1594	CTTCCACCCCAAGTGGGGCCATTAATAATACGATTCTATATGCGACGGGCTACATACAGTGGC	1653
Db	196	CTTCTACTCTCAAAATGGGGCCATTAGAAATCTACATTTGGCTACGAGGGCGCATATAGCTCAGC	255
QY	1654	ATTCTCTTGTGTGCTACTGGGAAGTACGGGAACCAAGTGTGAGAGACACACCTTGGGCAC	1713
Db	256	GTTCTCTGTTGTGTGTCTTCCGCCAATATACGAGAACATGTTGGGGAACACACACTGTGGCC	315
QY	1714	GGTCCACACCCACAGCGGCCCACTTCAAGGTGATCTGTGATGTAGACAGCACTGGACAATG	1773
Db	316	CGTCCAAATATGACAGTCTCACTACACAACTGGATTCGGGAGCTGTGTGAGACTGGCAAACTG	375
QY	1774	GGTCTGGGCGAGGATATGTCTTGTCCCAATGAGCT	1810
Db	376	GGCTGTGCTGATGATATATCCATTTTGTCTACACAGCT	412

```

RESULT 14
US-09-918-995-35960
: Sequence 35960, Application US/09918995
: Publication No. US20030073623A1
: GENERAL INFORMATION:
: APPLICANT: Hyseq, Inc.
: TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
: FILE REFERENCE: 20411-756
: CURRENT APPLICATION NUMBER: US/09/918_995
: CURRENT FILING DATE: 2001-07-30
: PRIOR APPLICATION NUMBER: US/09/235_076
: PRIOR FILING DATE: 1999-01-20
: NUMBER OF SEQ ID NOS: 38054
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 35960
: LENGTH: 405
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-918-995-35960

Query Match          5.2%; Score 209.8; DB 9; Length 405;
Best Local Similarity 72.7%; Pred. No. 3,9e-52;
Matches 271; Conservative 0; Mismatches 102; Indels 0; Gaps 0.

QY 2074 GGAGGAGGAGCCAGTAGACGCGTTTCAATCGAATGACCCCTTGCGCCCACTGT 2133
DB 13 GGAGGAGGAGTACACAGACGAGTAGCATATACACCGAATGACATCTGACACCCACAGT 72
QY 2134 GCATTTCACTGACTTCATCAACAAATGAGACCATTCTGGAAAGATTTGTGGCTGGGT 2193
DB 73 TACCTTGGTCACTTCATCAACAAATGAAACCCCTTAGAGAGAGATCTGTGGCTTGGGT 132
QY 2194 GACAGCTGGTTTCTGCAATATCCCATGACAGGAGCATTCCTAACAAGTAGACTGTGG 2253
DB 133 CACAGCACCTTCTGCAATATCCCATGCGAGGACATCCCAACACAGTAGACTGTGA 192
QY 2254 GAACGCGCTGGGGCTTTCCTCGACCCATATAACTCTTTGAGAGAACCCCTCTTGA 2313
DB 193 GAACAGAGTTGGCTTTCTGCTCGACCCATATAACTCTTTGATGAGAGAACCCCTCATCTT 252
QY 2314 CTCGCGCACTCCATCTACTTCCGAGGGAGCACAGATGCTGGGGCTGCGAGGTCAACCC 2373
DB 253 CTCGCCCTGGCAGATGTACTATTGAGAAAGGGCCAGATGCTGGGGCTCTGCAGCATCAATCC 312
QY 2374 CCTAGCTTGCTCCCAAGGCTCTGCTGTGCCCCCGAGACTCCCTGCTCTTCCACAGG 2433
DB 313 TGTGGCTGCTGCCCGACCTGGACAGCTGTGTGCCGGAGTACCCCTTCTTACCA 372
QY 2434 GGGCTTCTCTCAC 2446
DB 373 CGGCTTCTACTAC 385

```

```

RESULT 15
US-09-960-352-3010
: Sequence 3010, Application US/09960352
: Patent No. US20020137139A1
: GENERAL INFORMATION:
: APPLICANT: Warren, Wesley C.
: APPLICANT: Tao, Nengding
: APPLICANT: Byatt, John C.
: APPLICANT: Mathalagan, Nagappan
: TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION
: TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
: FILE REFERENCE: 16511.006/37-21(102981C
: CURRENT APPLICATION NUMBER: US/09/960.352
: CURRENT FILING DATE: 2001-09-24
: NUMBER OF SEQ ID NOS: 15112
: SEQ ID NO 3010
: LENGTH: 254
: TYPE: DNA
: ORGANISM: Bos taurus
: OTHER INFORMATION: Clone ID: 13-LIB34-066-Q1-E1-D1
US-09-960-352-3010

```

	Query Match	Score	DB	Length
Best Local Similarity	82.7%	Pred. No. 2e-44		
Matches	210	Conservative	0	Mismatches 44
				Indels 0
				Gaps 0
QY	1164	GGACTTCTCCTTTGGCCTCGAGCATTCAGTGGCCCAAGATCTTTGACGTTCGCTTCC	1223	
Db	1	GGACTTCTCCTTTGGCCTCGAGCGTTTCAGTGGCTTCTTGGACGTTCATTTCC	60	
QY	1224	AAGGAGAAAGACTAGTTATGAGATTAAGCCTCCAGAGAGCGCTTGGCCACTATGTGTGAA	1283	
Db	61	AGGGAGAAAGCGCTGGCTTATGAGATCAGCCTCAAGAGCGCTGTGTCTGTACGGTGGGA	120	
QY	1284	ATTCCCAGACACATGACAGACCCCTAGTGGATGAGAGCGTTTGGCAATGGGCAAGTAA	1343	
Db	121	ATATCTCCACACCAATGCTCATTCTCGTATATGATATGTGGCTTTGGCAATGGGTTACTTGG	180	
QY	1344	CCAGCGCCCTGACCCGTGGGGTGGATGCGCCCTACTTTGGCCACATGATGGATGGCACT	1403	
Db	181	CCACACCCCTGATTTGGTGGCGTATGAGTGCCTTACTCTGGCCACCTCATGAGACTGGCACT	240	
QY	1404	TGCTTTTGGAGTCC	1417	
Db	241	TGCTTTTGGAGTCC	254	

Search completed: May 20, 2003, 11:30:46
Job time : 834 secs

Tue May 20 13:18:47 2003

us-10-081-408-1.rst

Page 1

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2003, 03:28:53 ; Search time 4937 Seconds
(without alignments)
13252.940 Million cell updates/sec

Title: US-10-081-408-1

Perfect score: 4040
Sequence: 1 gtcctccaccacctatgc.....ccccaaaaaaaaaaaaa 4040

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:*
1: em_estda:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4008.2	99.2	4064	11	BC036368 Homo sapi
2	932.2	23.1	961	9	AL540526
3	817.4	20.2	929	14	BQ706857 AGENCOURT
4	801.2	19.8	887	13	B1818374 603032815
5	780	19.3	1039	9	AL568455
6	751.8	18.6	957	14	BQ878960 AGENCOURT

7	720.4	17.8	764	13	B1489920
8	675.2	16.7	767	12	BC033106
9	634	15.6	644	14	BM573838
10	631.4	15.6	985	14	BQ712543
11	622.4	15.4	966	14	BQ720486
12	608.6	15.1	694	12	BG718238
13	587	14.5	622	14	BM992567
14	567	14.0	1270	14	U25753
15	551.8	13.7	788	13	B1818437
16	549	13.6	703	12	BG721874
17	535.8	13.3	578	13	BM665550
18	527.8	13.1	531	10	BE673388
19	525	13.0	526	14	BM831002
20	508	12.6	655	10	AM833957
21	506	12.5	506	9	A1279210
22	493	12.2	493	9	AL047423
23	482	11.9	482	9	A1129246
24	480.4	11.9	484	9	A1082117
25	469.6	11.6	641	10	BB636555
26	434.4	10.8	436	9	A1042036
27	431	10.7	431	14	BM831093
28	420.8	10.4	570	13	B1681765
29	414.2	10.3	556	13	BM483607
30	413.8	10.2	417	9	AA807904
31	407.8	10.1	412	12	BF433374
32	407	10.1	431	9	A1268118
33	399.2	9.9	541	14	H42560
34	397.2	9.8	424	10	AM131121
35	393.4	9.7	476	9	A1807879
36	384.6	9.5	511	12	BE758045
37	382.2	9.5	410	9	A1871588
38	377.2	9.3	465	9	AA036974
39	376.4	9.3	381	9	A1458535
40	376.4	9.3	426	13	B1013968
41	362.6	9.0	511	14	H43038
42	354.2	8.8	359	10	BE048629
43	354	8.5	354	14	BM823391
44	342.8	8.5	502	10	BE479569
45	333.6	8.3	407	14	W24606

ALIGNMENTS

RESULT 1
BC036368
LOCUS
DEFINITION Homo sapiens, amine oxidase, copper containing 3 (vascular adhesion protein 1), clone IMAGE:4826423, mRNA.
ACCESSION BC036368
VERSION BC036368.1 GI:22137778
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 4064)
Strausberg, R.
Direct Submission
Submitted (05-AUG-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NMGRI) & Shitaki Toshitsuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadane@systemsbiology.org

REMARK COMMENT

Anup Madan, Jessica Fahey, Erin Helton, Mark Keltman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAX Plate: 33 Row: n Column: 3
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA 91: 6806883
This clone has the following problem: frame shifted.

FEATURES

source

Location/Qualifiers

1:4064

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4826423"

/rissue_type="testis"

/clone_id="NIH_MGC_97"

/lab_host="DH105"

/note="Vector: pBluescript"

BASE COUNT 860 a 1177 c 1097 g 930 t

ORIGIN

Query Match 99.2%; Score 4008.2; DB 11; Length 4064;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 4021; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```
OY 16 AGTCCAGGATCTGACTACCGGGAACCTCAGCCAGAGTCGGGAGGCCGCCACCCGCTC 75
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6 AGTCACAGGATCTGACTACCGGGAACCTCAGCCAGAGTCGGGAGGCCGCCACCCGCTC 65

OY 76 CAGGAGCCACAGAGAGCCCGCTCTGCTGCTGAGATACATCTCTCTCTCTCTCTCTCT 135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 66 CAGGAGCCACAGAGAGCCCGCTCTGCTGCTGAGATACATCTCTCTCTCTCTCTCTCT 125

OY 136 ATCAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 126 ATCAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 185

OY 196 GCGCGTATACCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 186 GCGCGTATACCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 245

OY 256 GGGTGAACCCAGAGCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 246 GGGTGAACCCAGAGCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 305

OY 316 ACACCCCTGGCCAGAGAGCTGTTGACAGACCTGACCCGAGAGAGAGAGAGAGAGAG 375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 306 ACACCCCTGGCCAGAGAGCTGTTGACAGACCTGACCCGAGAGAGAGAGAGAGAGAG 365

OY 376 GCGCTTCTGACCCAGAGGCTGGGGCCAGGGCTGAGATGACCCGAGCCGCGCTC 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 366 GCGCTTCTGACCCAGAGGCTGGGGCCAGGGCTGAGATGACCCGAGCCGCGCTC 425

OY 436 GGACAACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 495
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 426 GGACAACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 485

OY 496 CTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 555
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 486 CTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 545

OY 556 GCAACCCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 615
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 546 GCAACCCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 605

OY 616 GCGGAGCTGACTGTGAGAGCTATGAGAGCCCTGCGCTATACCGAGCCCGCGTCT 675
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 606 GCGGAGCTGACTGTGAGAGCTATGAGAGCCCTGCGCTATACCGAGCCCGCGTCT 665

OY 676 GTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 735
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 666 GTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 725
```

```
OY 736 TGGGCTTCTCCACCACTGTTGCTTACAAGCACCGGGGAGGAGAACTGTGACAAATGAC 795
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 726 TGGGCTTCTCCACCACTGTTGCTTACAAGCACCGGGGAGGAGAACTGTGACAAATGAC 785

OY 796 CAGGCTCCCGGTGCTGCAATCAGGGAGCCGGCCACCTGTTGGCTCTACTACAA 855
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 786 CAGGCTCCCGGTGCTGCAATCAGGGAGCCGGCCACCTGTTGGCTCTACTACAA 845

OY 856 CATCTGGGGGGGTGGTCTCTCTGACACAGTGGGCTTGGAGCTGCTACTGAAACCA 915
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 846 CATCTGGGGGGGTGGTCTCTCTGACACAGTGGGCTTGGAGCTGCTACTGAAACCA 905

OY 916 GGGCTTGAACCTTGGCCGCTGAGATATCCAGAGGTTTCTATCAAGGCCGCTACTACA 975
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 906 GGGCTTGAACCTTGGCCGCTGAGATATCCAGAGGTTTCTATCAAGGCCGCTACTACA 965

OY 976 CAGCTGGCCAGCTGAGAGCCAGTTTGAAGCCGCTGCTGATGATGCTGATCC 1035
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 966 CAGCTGGCCAGCTGAGAGCCAGTTTGAAGCCGCTGCTGATGATGCTGATGATCC 1025

OY 1036 AGACAAATGGGACAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1095
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1026 AGACAAATGGGACAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1085

OY 1096 CCTCTACAGTTCTATCCCAAGGCCGCTTCAAGTGTCCAGAGGAGTGAAGTGGCTC 1155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1086 CCTCTACAGTTCTATCCCAAGGCCGCTTCAAGTGTCCAGAGGAGTGAAGTGGCTC 1145

OY 1156 CTCACGTGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1146 CTCACGTGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1205

OY 1216 TCGCTTCCAAAGGAAAGACTATGTTATGAGTAAAGCTTCCAAAGGCTTGGCATCTA 1275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1206 TCGCTTCCAAAGGAAAGACTATGTTATGAGTAAAGCTTCCAAAGGCTTGGCATCTA 1265

OY 1276 TGGTGAATTTCCGACAGCAATAGAGACCCGCTATGAGTGAAGGCTTGGCATGGG 1335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1266 TGGTGAATTTCCGACAGCAATAGAGACCCGCTATGAGTGAAGGCTTGGCATGGG 1325

OY 1336 CAAGTACACAGCCGCTGACCCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1395
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1326 CAAGTACACAGCCGCTGACCCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1385

OY 1396 CTGGACATCTCTTGGAGTCCGAGGCCCAAGCAATAGATGATCTTTGTGTT 1455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1386 CTGGACATCTCTTGGAGTCCGAGGCCCAAGCAATAGATGATCTTTGTGTT 1445

OY 1456 TGAACAAACAGAGGCTTCCCTGCGGGGAGACACACTGATCTACTGCACTACTT 1515
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1446 TGAACAAACAGAGGCTTCCCTGCGGGGAGACACACTGATCTACTGCACTACTT 1505

OY 1516 TGGGGTCTTGGGAAACGGTGTGCTGCTGATGATCTATGCTCACTTGTCTCAATGTA 1575
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1506 TGGGGTCTTGGGAAACGGTGTGCTGCTGATGATCTATGCTCACTTGTCTCAATGTA 1565

OY 1576 CTATGTGGGATACGGTCTCCACCCAGTGGGGCATAGAAATAGATTCTATGTCAC 1635
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1566 CTATGTGGGATACGGTCTCCACCCAGTGGGGCATAGAAATAGATTCTATGTCAC 1624

OY 1636 GGGCTATACAGCTCGGCACTCTCTTGGGTGCTACTGAGGAAATGAGGAAACAAAGTGC 1695
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1626 GGGCTATACAGCTCGGCACTCTCTTGGGTGCTACTGAGGAAATGAGGAAACAAAGTGC 1614

OY 1696 AGACACACCCCTGGGCAAGGCTCACACCCAGCCCACTTCAAGGTGATCTGATGT 1755
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1686 AGACACACCCCTGGGCAAGGCTCACACCCAGCCCACTTCAAGGTGATCTGATGT 1744

OY 1756 AGCAGAGCTGGAGAACTGGGCTGGGCGAGAGATATGTTCTTGTCCCAATGGCTGAGCC 1815
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1746 AGCAGAGCTGGAGAACTGGGCTGGGCGAGAGATATGTTCTTGTCCCAATGGCTGAGCC 1804

OY 1816 CTGGAGCCCTGAGCACCAGCTGACAGAGCTTGCAGTGAACCCGGAAGCTGCTGAGATGGA 1875
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

[illegible]

Db	2885	CAGGTCTCTTCCCTTTCGTGCTTCCCTCTCTCTACACCTACTTCTCCTCCTCCTCCTGTTCC	2944
Qy	2956	TGCCCTCTCTTCAATCCCTGCAATTTTCTCCGAATCTCGAGGGGAAATATCCATATGTCCAG	3015
Db	2945	TGCCCTCTCTTCAATCCCTGCAATTTTCTCCGAATCTCGAGGGGAAATATCCCTATGTCCAG	3004
Qy	3016	CCCCGTGACCCCCCGAGCCCTCAGTTTTCAGTCAGATTCCGTCCTCTCTCCAGCCCAT	3075
Db	3005	CCCCGTGACTCCCCCGAGCCCTCAGTTTTCAGTCAGATTCCGTCCTCTCTCCAGCCCAT	3064
Qy	3076	GGAAGTCTCAAGGTCAAGGGGACCCCTAATCAGAGTGGCCAAATCCCTGTGTCTTCCCT	3135
Db	3065	GGAAGTCTCAAGGTCAAGGGGACCCCTAATCAGAGTGGCCAAATCCCTGTGTCTTCCCT	3124
Qy	3136	TGTGTCTGTGCTTATATGGAGATGAGAGTTGCTTCAACCCCTGTCTCTGGGGCTGGGTGTG	3195
Db	3125	TGTGTCTGTGCTTATATGGAGATGAGAGTTGCTTCAACCCCTGTCTCTGGGGCTGGGTGTG	3184
Qy	3196	TTTACAGACAGCTGCTCTGTGATTTGTGTGCTGCTGCTCATGCTCTCTATACAGAG	3255
Db	3185	TTTACAGACAGCTGCTCTGTGATTTGTGTGCTGCTGCTCATGCTCTCTATACAGAG	3244
Qy	3256	GATGCTCATCTGTACAGCAGCAGCTCAAGTTAGCATTTCAAGTATTTGGGGTGCATG	3315
Db	3245	GATGCTCATCTGTACAGCAGCAGCTCAAGTTAGCATTTCAAGTATTTGGGGTGCATG	3304
Qy	3316	ATAATGAAGATGGCCATTTTGTACAGGGGCTGTATTCTGCACAGCCTGTTTGGAG	3375
Db	3305	ATAATGAAGATGGCCATTTTGTACAGGGGCTGTATTCTGCACAGCCTGTTTGGAG	3366
Qy	3376	GCTGAGTGGAAACAAAGGGTGGGCATCAAAAGATGAAAGCCAAAGCCCTTCAACTCA	3435
Db	3365	GCTGAGTGGAAACAAAGGGTGGGCATCAAAAGATGAAAGCCAAAGCCCTTCAACTCA	3424
Qy	3436	GCCACCCAGCCAGGAGGGGCTGTCCATCATCTTACAGGCAATGGCAATGAGCTAGGGCCGTG	3495
Db	3425	GCCACCCAGCCAGGAGGGGCTGTCCATCATCTTACAGGCAATGGCAATGAGCTAGGGCCGTG	3488
Qy	3496	GGTGAGGTGGGGCTGTGGCCTAGTGGGAGGGGCTGGCCTGGGTGGGACAGGGCTGGC	3555
Db	3485	GGTGAGGTGGGGCTGTGGCCTAGTGGGAGGGGCTGGCCTGGGTGGGACAGGGCTGGC	3544
Qy	3556	CTGGCCAGGCTTGGGCTCCATTCCTCACTACCTGCTTCCCTCTGAGGCTGGATTTGGGG	3615
Db	3545	CTGGCCAGGCTTGGGCTCCATTCCTCACTACCTGCTTCCCTCTGAGGCTGGATTTGGGG	3604
Qy	3616	ATGGGACAAAGAAATATGCAAGAGATGAGAAACACAGAACTTTTCTCTAAAGGACT	3675
Db	3605	ATGGGACAAAGAAATATGCAAGAGATGAGAAACACAGAACTTTTCTCTAAAGGACT	3664
Qy	3676	GCTTAATCAATTGTGATACAGCTTCAATACATATGATGACGTAAAAAATAATGT	3735
Db	3665	GCTTAATCAATTGTGATACAGCTTCAATACATATGATGACGTAAAAAATAATGT	3724
Qy	3736	ATGCTTTATATTAATATATGTAATACTTCAAGGTAAAAAGCAAGCACAAGAAATGT	3795
Db	3725	ATGCTTTATATTAATATATGTAATACTTCAAGGTAAAAAGCAAGCACAAGAAATGT	3788
Qy	3796	GTAATAGCGCACTTCCATTGTGTGTTTCAGAAAGAGATAGATATTAACACATATTTGCTT	3855
Db	3785	GTAATAGCGCACTTCCATTGTGTGTTTCAGAAAGAGATAGATATTAACACATATTTGCTT	3844
Qy	3856	ATGATAGCTTATAGAAATTAAGGGTAACACGTACTTCTTTGGAGGGGAACAAGTAG	3915
Db	3845	ATGATAGCTTATAGAAATTAAGGGTAACACGTACTTCTTTGGAGGGGAACAAGTAG	3904
Qy	3916	GTTAGAGCAGAGAGAGGAGGCTTAACTTAACCTTTTGATCAATTTTGATATTTT	3975
Db	3905	GTTAGAGCAGAGAGGAGGCTTAACTTAACCTTTTGATCAATTTTGATATTTT	3966
Qy	3976	GAAACATGTGACTGTATTACCTATTCAAAATTAACAATAATATGGGCCAAAAAATAAATA	4035
Db	3965	GAAACATGTGACTGTATTACCTATTCAAAATTAACAATAATATGGGCCAAAAAATAAATA	4024

Query Match	Best Local Similarity	Matches	Score	DB	Length
1457	GAACGAAACGAGGGCTCCCTCGGGGACACCACTCGAGCTCTACTCGCACTACTTT	1516	20.2%; 97.1%;	271 c	269 g
14	GAGGGAACACGAGGCTCCCTCGGGGACACCACTCGAGCTCTACTCGCACTACTTT	73	20.2%; 97.1%;	271 c	269 g
1517	GGGGCTCTTGCAGGAACGGTCTGGTTCGTCACATCATATGTCACACCTTGTCTCACTATGAC	1576	20.2%; 97.1%;	271 c	269 g
74	GGGGCTCTTGCAGGAACGGTCTGGTTCGTCACATCATATGTCACACCTTGTCTCACTATGAC	133	20.2%; 97.1%;	271 c	269 g
1577	TATGTGTGGGATAGGCTCTTCCACCCCACTGGGGCCATGAAATACGATTTATGACACG	1636	20.2%; 97.1%;	271 c	269 g
134	TATGTGTGGGATAGGCTCTTCCACCCCACTGGGGCCATGAAATACGATTTATGACACG	193	20.2%; 97.1%;	271 c	269 g
1637	GGCTCATCATAGTGGGCACTTCTCTTGGTGTACTGGGAAGTAGGGGAACCAAGTGTCA	1696	20.2%; 97.1%;	271 c	269 g
194	GGCTCATCATAGTGGGCACTTCTCTTGGTGTACTGGGAAGTAGGGGAACCAAGTGTCA	253	20.2%; 97.1%;	271 c	269 g
1697	GAGCACACCCCTGGGACGGTCTCCACACCCACAGCGCCCACTTGCAAGGTGATCTGATGTA	1756	20.2%; 97.1%;	271 c	269 g
254	GAGCACACCCCTGGGACGGTCTCCACACCCACAGCGCCCACTTGCAAGGTGATCTGATGTA	313	20.2%; 97.1%;	271 c	269 g
1757	GCAGGACTGTGAGACATGCGGTCTGGGCGGAGATATGCTTTTGTCCCATATGCTGTGCC	1816	20.2%; 97.1%;	271 c	269 g
314	GCAGGACTGTGAGACATGCGGTCTGGGCGGAGATATGCTTTTGTCCCATATGCTGTGCC	373	20.2%; 97.1%;	271 c	269 g
1817	TGGAGCCCTGACACACAGCTGTGCAGAGCGTCGACAGGTACCCGGAAAGTGTGAGATGAG	1876	20.2%; 97.1%;	271 c	269 g
374	TGGAGCCCTGACACACAGCTGTGCAGAGCGTCGACAGGTACCCGGAAAGTGTGAGATGAG	433	20.2%; 97.1%;	271 c	269 g
1877	GAGCAGGCGGCTTCTCTGTGGGAACGCGACCCCTCGCTTACCTTACCTGGGACAGCAAC	1936	20.2%; 97.1%;	271 c	269 g
434	GAGCAGGCGGCTTCTCTGTGGGAACGCGACCCCTCGCTTACCTTACCTGGGACAGCAAC	493	20.2%; 97.1%;	271 c	269 g
1937	CACAGCAACAGTGGGGTACCCCGGGGGCTACCGCATCCAGATGCTCAGCTTGTGTGA	1996	20.2%; 97.1%;	271 c	269 g
494	CACAGCAACAGTGGGGTACCCCGGGGGCTACCGCATCCAGATGCTCAGCTTGTGTGA	553	20.2%; 97.1%;	271 c	269 g
1997	GAGCGGCTGCCCCAAACAGCTCATGCGGAGAGGCTTACGCTGGGAGAGGTACCAAGCTG	2056	20.2%; 97.1%;	271 c	269 g
554	GAGCGGCTGCCCCAAACAGCTCATGCGGAGAGGCTTACGCTGGGAGAGGTACCAAGCTG	613	20.2%; 97.1%;	271 c	269 g
2057	GCGTGAACCCAGGGAAGAGGAGGAGGCCAGTAGAGAGCAGCTTTTCATCGAATGAC	2116	20.2%; 97.1%;	271 c	269 g
614	GCGTGAACCCAGGGAAGAGGAGGAGGCCAGTAGAGAGCAGCTTTTCATCGAATGAC	673	20.2%; 97.1%;	271 c	269 g
2117	CCTTGGGCCCCCACTGTGATTTCACTGATTCATCAACATAGACCATTTGCTGGAAG	2176	20.2%; 97.1%;	271 c	269 g
674	CCTTGGGCCCCCACTGTGATTTCACTGATTCATCAACATAGACCATTTGCTGGAAG	733	20.2%; 97.1%;	271 c	269 g
2177	GATTTGTGGCTGGGTGACAGCTGTTTTCGATATATCCCATGACGAGGACATTTCT	2236	20.2%; 97.1%;	271 c	269 g
734	GATTTGTGGCTGGGTGACAGCTGTTTTCGATATATCCCATGACGAGGACATTTCT	793	20.2%; 97.1%;	271 c	269 g
2237	AACACAGTACATGTTGGGGA - CGGGGTGGGCTTCTTCCGACACCTAATCTCTT -	2293	20.2%; 97.1%;	271 c	269 g
794	AACACAGTACATGTTGGGGA - CGGGGTGGGCTTCTTCCGACACCTAATCTCTT	853	20.2%; 97.1%;	271 c	269 g

QY	2294	-GAGCAACACCCCTCTCTACTGCGCG-ACGCATCTACTTCCGAGGG	2342
Db	854	GACGAAAGACCCCTCTCTACTGCGCGACACTCATCTACTTCCAGGG	904
RESULT 4			
LOCUS	B1818374	887 bp	RNA linear EST 04-OCT-2001
DEFINITION	603032815F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5174187 5',		
ACCESSION	B1818374		
VERSION	B1818374.1	GI:15929130	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.		
AUTHORS	1 (bases 1 to 887)		
TITLE	NIH-MGC http://imgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabs-remail.nih.gov		
	Tissue Procurement: Life Technologies, Inc.		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://imgc.lnl.gov		
	Plate: LNL11433	row: m	column: 04
	High quality sequence stop: 882.		
FEATURES	Location/Qualifiers		
source	1..887		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:5174187"		
	/clone_11b="NIH_MGC_115"		
	/lab_host="DH10B"		
	/note="Organ: pooled brain, lung, testis; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH_MGC Library."		
BASE COUNT	252 a 170 c 238 g 227 t		
ORIGIN			
Query Match	19.8%; Score 801.2; DB 13; Length 887;		
Best Local Similarity	98.8%; Pred. No. 2e-144;		
Matches	881; Conservative 0; Mismatches 3; Indels 8; Gaps 7.		
QY	3152	TGGAGTAGAGAGTGTCTCTACCCCTGTCTGGGGCTGGTGTCTTTCAGACAGCTGCT	3211
Db	1	TGGAGTAGAGAGTGTCTCTACCCCTGTCTGGGGCTGGTGTCTTTCAGACAGCTGCT	60
QY	3212	TCTGTGATTTGTGTGCTGCTGCTCATGCTCTCTTAGAGGAGATGATCATCGTACA	3271
Db	61	TCTGTGATTTGTGTGCTGCTGCTCATGCTCTCTTAGAGGAGATGATCATCGTACA	120
QY	3272	GCAGCAGCTCAAGTATGACATTTCAAGTATTTGGGGGTGCAATGATTAATGAAGATGCC	3331
Db	121	GCAGCAGCTCAAGTATGACATTTCAAGTATTTGGGGGTGCAATGATTAATGAAGATGCC	180
QY	3332	ATTTTGTACAGGGGCTGTATCTTCACACAGCTTTTGGAGAGCTGGAGTCGAAACA	3391
Db	181	A-TTTGTACAGGGGCTGTATCTTCACACAGCTTTTGGAGAGCTGGAGTCGAAACA	239

```
OY 3392 AGGTTGGGATCAAAAGATGAGAAAGCCAAAGCCCTACAACTCAGCCAGCCAGCCAGAG 3451
|||||
Db 240 AGGTTGGGATCAAAAGATGAGAAAGCCAAAGCCCTACAACTCAGCCAGCCAGCCAGAG 299
OY 3452 GGGCTGTCCAACTCACTTACAGGATGCGGAATGAGCTGGGCTGGTGGAGTGGGGTCT 3511
|||||
Db 300 GGGCTGTCCAACTCACTTACAGGATGCGGAATGAGCTGGGCTGGTGGAGTGGGGTCT 359
OY 3512 GGCTATGAGGAGGGGCTGGCTGGTGGGAGAGGCTGGGCTGGTCCAGGCTTGG 3571
|||||
Db 360 GGCTATGAGGAGGGGCTGGCTGGTGGGAGAGGCTGGGCTGGTCCAGGCTTGG 418
OY 3572 CTCATTCCTCATCAGCTGCTGCTCCCTCCTGAGTGTGATTTGGGAGCAAGAAAT 3631
|||||
Db 419 CTCATTCCTCATCAGCTGCTGCTCCCTCCTGAGTGTGATTTGGGAGCAAGAAAT 478
OY 3632 AGCAAGAGATGAGAAACACAGAAAC-TTTTCTCTTAAGGAGCTGTTAAATCAATCT 3650
|||||
Db 479 AGCAAGAGATGAGAAACACAGAAAC-TTTTCTCTTAAGGAGCTGTTAAATCAATCT 538
OY 3691 GATACAGCTTACAAATCAATAGATGAGCTAAATAATATGATGCTTTATATCT 3750
|||||
Db 539 GATACAGCTTACAAATCAATAGATGAGCTAAATAATATGATGCTTTATATCT 598
OY 3751 AATATGATTAATCTTCAAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3810
|||||
Db 599 AATATGATTAATCTTCAAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 658
OY 3811 CATTTGCTTCAAGAGGAGTAAATATATATATATATATATATATATATATATATAT 3870
|||||
Db 659 CATTTGCTTCAAGAGGAGTAAATATATATATATATATATATATATATATATATAT 718
OY 3871 GAATTAATGGGTAACTGATTAATTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3950
|||||
Db 719 GAATTAATGGGT-AACTGATTAATTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 777
OY 3931 GGGAA-GGTTCTTACACTTACACCC-TTTTGTACATTTTGAATTTGAACATGTACT 3988
|||||
Db 778 GGGAAAGGGGCTTAACTTACACCC-TTTTGTACATTTTGAATTTGAACATGTACT 837
OY 3989 GTATTACTTTCATAAATAAATAGGCCCAAAAAAAAAAAAAAAAAAAAA 4040
|||||
Db 838 GTATTACTTATCAAA--TAAATAAATAGGGCCCAAAAAAAAAAAAAA 887

RESULT 5
AL568455 1039 bp mRNA linear EST 16-FEB-2001
LOCUS AL568455
DEFINITION AL568455 L1T-FL002_P1 Homo sapiens cDNA clone CS0DE001YN01 3 prime
, mRNA sequence.
ACCESSION AL568455
VERSION AL568455
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 1039)
L1.W.B., Gruber,C., Jesse,J. and Polyes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..1039
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DE001YN01"
/clone_1lb="L1T-FL002_P1"
/lab_host="DH10B"
/note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand
```

```
cdna was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cdna was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
f.liang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 320 a 259 c 188 g 239 t 33 others
ORIGIN

Query Match 19.3% Score 780; DB 9; Length 1039;
Best Local Similarity 85.9%; Pred. No. 2.5e-140;
Matches 887; Conservative 33; Mismatches 89; Indels. 23; Gaps 5;

OY 2995 GGGGATATCCCTATGCTCCAGCCCTGTAAGTCCCTCAGCTTTCAGTCAAGTT 3054
|||||
Db 1032 RGGGATATCCCTATGCTCCAGCCCTGTAAGTCCCTCAGCTTTCAGTCAAGTT 973
OY 3055 CCGTCTCCTCCAGCCCTATGAGATCTCAAGTCAAGGAGGAGGAGGAGGAGGAGGAGG 3114
|||||
Db 972 CCGTCTCCTCCAGCCCTATGAGATCTCAAGTCAAGGAGGAGGAGGAGGAGGAGGAGG 913
OY 3115 AATCCCTGCTGCTGCTCCCTTGTCTGCTGCTTATTTGGAGTATGCTGCTTACC 3174
|||||
Db 912 AATCCCTGCTGCTGCTCCCTTGTCTGCTGCTTATTTGGAGTATGCTGCTTACC 853
OY 3175 CCTGCTCCTGGGGGCTGGTGTGCTTCAAGACAGCTGCTGCTGCTGCTGCTGCTG 3224
|||||
Db 852 CCTGCTCCTGGGGGCTGGTGTGCTTCAAGACAGCTGCTGCTGCTGCTGCTGCTGCTG 793
OY 3225 CTCATTCCTCTATAGAGAGATGCTATGCTATACAGACAGCTCAAGTTACATTTTC 3294
|||||
Db 792 CTCATTCCTCTATAGAGAGATGCTATGCTATACAGACAGCTCAAGTTACATTTTC 733
OY 3295 AAGTATTTGGGGGTGCAATGATTAAGATGAGGATTTTGTACAGAGGCTGTGATT 3354
|||||
Db 732 AAGTATTTGGGGGTGCAATGATTAAGATGAGGATTTTGTACAGAGGCTGTGATT 673
OY 3355 CTCGACAGCCTGTTTGGAGGCTGAGAGTGAACAAAGGAGGAGGAGGAGGAGGAGG 3414
|||||
Db 672 CTCGACAGCCTGTTTGGAGGCTGAGAGTGAACAAAGGAGGAGGAGGAGGAGGAGG 613
OY 3415 GCCAAGGCCCTTACAACTCCAGCCAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3474
|||||
Db 612 GCCAAGGCCCTTACAACTCCAGCCAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 553
OY 3475 ATGCGAATGAGCTGGGCTGAGTGGAGTGGGCTGAGGCTAGTGGGAGAGGAGGCTG 3534
|||||
Db 552 ATGCGAATGAGCTGGGCTGAGTGGAGTGGGCTGAGGCTAGTGGGAGAGGAGGCTG 493
OY 3535 CTGGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3594
|||||
Db 492 STGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 433
OY 3595 CTCCTGAGGCTGATTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3654
|||||
Db 432 YTCCTGAGGCTGATTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 373
OY 3655 AACTTTTCTCTAAGAGCTGTTAAATCAATCTGATACAGGCTTACATATAGT 3714
|||||
Db 372 TWTTTTCTCTAAGAGCTGTTAAATCAATCTGATACAGGCTTACATATAGT 313
OY 3715 ATGCAAGTAAATAATTTGATGCTTTATATATATATATATATATATATATATATAT 3774
|||||
Db 312 ATGCAAGTAAATAATTTGATGCTTTATATATATATATATATATATATATATATATAT 253
OY 3775 AAAGGCAAGCCACA-----GAAATGTATATAGCCAGCTTCCATTTGTGTTCA 3826
|||||
Db 252 AAAGGCAAGCTATCAGAAATGTTATATAGCTTCCATTTGTGTTCA 193
OY 3827 AGG-AGTAGAATATTAACATATA-----TTGCTATATATATATATATATATATAT 3878
```

```

Db 192 AGGTGAGTAAATATAAATTAATTCCTTATGTTATGCTCTTATCTAGAAATTAAT 133
Oy 3879 GGGTAACACT--GATTACTTTGGGAGGGAACAGTAGCTGAGACAGAGAGGANG 3936
Db 132 GGGTAACACTAGATTATTTTGGGAGGGAACAGTAGCTGAGACAGAGAGGANG 73
Oy 3937 GGTCTT-----AACACTTACACCCCTTTGTACATTTTGAATTTGAACCTGTGACTGA 3991
Db 72 GTTTTGTGCACTACTTCTCTTTTGTACATTTTATTTTGAATATATGATTKTA 13
Oy 3992 TTACCTATTCAA 4003
Db 12 TTTTATATTAA 1

RESULT 6
B0878960 957 bp mRNA linear EST 16-AUG-2002
LOCUS AGENCOURT_8231141 lupskl_dorsal_root_ganglion Homo sapiens cDNA
DEFINITION clone IMAGE:6182295 5', mRNA sequence.
ACCESSION B0878960.1 GI:22270968
VERSION B0878960.1 GI:22270968
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 957)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgs@bhs-riemail.nih.gov
Tissue procurement: Dr. James R. Lupski
cDNA library preparation: Life Technologies, Inc.
cDNA library arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM13568 row: a column: 16
High quality sequence stop: 634.
Location/Qualifiers
1..957
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone IMAGE:6182295"
/clone_11b="Lupskl_dorsal_root_ganglion"
/sex="male"
/tissue_type="dorsal root ganglia"
/dev_stage="adult, 36 yr"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6 (Life Technologies); Site_1:
NcoI; Site_2: SalI; cDNA made by oligo-dT priming.
directionally cloned using the following adaptors:
5'-TCGACCAACGCGCTCG-3' and
5'-GACTGATTCATGATCGAGCGGAGCGGCGCT(15)-3'. Size selected >
1 kb for average insert length 1.7 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
BASE COUNT 168 a 315 c 279 g 194 t 1 others
ORIGIN
Query Match 18.6%; Score 751.8; DB 14; Length 957;
Best Local Similarity 92.5%; Pred. No. 6,8e-135;
Matches 825; Conservative 0; Mismatches 58; Indels 9; Gaps 3;
Oy 29 TGACTACCGGGAACCTCAGCCAGATCGCGGAGCCGCCACCCGCTCCAGGACCAACAG 88
Db 1 TGACTACCGGGAACCTCAGCCAGATCGCGGAGCCGCCACCCGCTCCAGGACCAACAG 60

```

```

Oy 89 ACCCCCCGCTTGTGCTGCGCTGAGAAATACATTCCTCTCTTGGTGAATACAGCTGTCCCT 148
Db 61 ACCCCCCGCTTGTGCTGCGCTGAGAAATACATTCCTCTCTTGGTGAATACAGCTGTCCCT 120
Oy 149 CTTCGTGGGAAATGAAACCAAGAAACATCTCTGCTCTCTTCTATCTTGGCCGTATACAC 208
Db 121 CTTCGTGGGAAATGAAACCAAGAAACATCTCTGCTCTCTTCTATCTTGGCCGTATACAC 180
Oy 209 ATCTTGGCTTGGTGTGCTCTCTGCTGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 268
Db 181 ATCTTGGCTTGGTGTGCTCTCTGCTGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240
Oy 269 CAGCTTCCCATTTGCTGCTCTCTCTATCTCTCCAGTCCAGCTCTGACACACCTGTGACG 328
Db 241 CAGCTTCCCATTTGCTGCTCTCTCTATCTCTCCAGTCCAGCTCTGACACACCTGTGACG 300
Oy 329 ACCCAGCTGTTTGCAGACCTGAGCCGAGAGAGAGCTGAGCTGTGATGCGCTTTTGACC 388
Db 301 ACCCAGCTGTTTGCAGACCTGAGCCGAGAGAGAGCTGAGCTGTGATGCGCTTTTGACC 360
Oy 389 CAGCGGCTGGGCGCAGAGGCTGTGTGATGACAGCCAGCCCGGCTCTGGACAACTGTGTC 448
Db 361 CAGCGGCTGGGCGCAGAGGCTGTGTGATGACAGCCAGCCCGGCTCTGGACAACTGTGTC 420
Oy 449 TTCTCAGTGAAGTTGAGCTGCTCTCCCAAGGCTGACAGCCCTGCTCACTTGGACAGAGG 508
Db 421 TTCTCAGTGAAGTTGAGCTGCTCTCCCAAGGCTGACAGCCCTGCTCACTTGGACAGAGG 480
Oy 509 AGCCCCCACTGCCCCGCGGAGAGCAGCTGCCATGCTCTTCTTGGCAGGCAACCCAGCCC 568
Db 481 AGCCCCCACTGCCCCGCGGAGAGCAGCTGCCATGCTCTTCTTGGCAGGCAACCCAGCCC 540
Oy 569 AACGTAGAGAGCTGTGTGCTGGGCGCAGCTGCTCCACCTCTCAATATGGGGAGCTGACT 628
Db 541 AACGTAGAGAGCTGTGTGCTGGGCGCAGCTGCTCCACCTCTCAATATGGGGAGCTGACT 600
Oy 629 GTGAGAGCTGATGAGAGGCCCCCTGCTATACAGCAGGCCCCGCTGTTCGAAGATAC 688
Db 601 GTGAGAGCTGATGAGAGGCCCCCTGCTATACAGCAGGCCCCGCTGTTCGAAGATAC 660
Oy 689 GTGAGATATGACCAATGATATTTCAACAGAGAGCTGCCAGCTTCTGGCTTCTCCAC 748
Db 661 GTGAGATATGACCAATGATATTTCAACAGAGAGCTGCCAGCTTCTGGCTTCTCCAC 720
Oy 749 CACTGTGCTTCAACAAGCAGCGGG---ACGGAACTGCTGTGATGACCAACG---CT 802
Db 721 CACTGTGCTTCAACAAGCAGCGGGAGCACTGGGTGCAATGACCAACGCGCTCC 780
Oy 803 CCCCGTGTCTGCAATCAGGAG---ACCGGCGCACTGTGTTGGCTCTACTACACATC 859
Db 781 CCCCGTGTCTGCAATCAGGAGAGCCCGGGCCACCTTGGNTTTGGCCCTACTACACAA 840
Oy 860 TCGGCGCTGGGCTTCTCTGCAACAGCTGGGCTTGAAGCTGCTAATCAAC 911
Db 841 AATCTCGGGCGGCTTGGGCTTCTCTGCGCCCAAGCGGGGCTTGGGAAC 892

RESULT 7
B1489920 764 bp mRNA linear EST 28-AUG-2001
LOCUS 603031674F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5172811 5',
DEFINITION mRNA sequence.
ACCESSION B1489920
VERSION B1489920.1 GI:15329148
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 764)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

```



```

|||||
Db 121 AAGGTACGGACCCCTAATCAGAGTGGCAATCCCTGTGTCTCTCCCTTGTCTGT 180
|||
Qy 3145 TCGTATTGGAGTAGAGTGTCTCTACCCCTGCTGCTGGGCTGTGTTTCAGGAC 3204
|||||
Db 181 TCGTATTGGAGTAGAGTGTCTCTACCCCTGCTGCTGGGCTGTGTTTCAGGAC 240
|||||
Qy 3205 AGCTGCTTGTGTCATTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3264
|||||
Db 241 ACTGCTCTGTGTCATTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
|||||
Qy 3265 CCGAGACAGACAGCTCAATGATGATGATGATGATGATGATGATGATGATGATG 3324
|||||
Db 301 CCGAGACAGACAGCTCAATGATGATGATGATGATGATGATGATGATGATGATG 360
|||||
Qy 3325 AATGCCATTTTGTACAGGCTCTGTATTCTGCAACAGCCTGTTGGAGGCTGGAGTG 3384
|||||
Db 361 AATGCCATTTTGTACCA-GGCTCTGTATTCTGCAACAGCCTGTTGGAGGCTGGAGTG 419
|||||
Qy 3385 GAACAAAGGCTGGCATCAAGATGAGAAAGCCCACTCAACTCCAGCCACCAG 3444
|||||
Db 420 GAACAAAGGCTGGCATCAAGATGAGAAAGCCCACTCAACTCCAGCCACCAG 479
|||||
Qy 3445 CCAAGAGGGGCTGCCAATCATTGATGATGATGATGATGATGATGATGATGATG 3504
|||||
Db 480 CCAAGAGGGGCTGCCAATCATTGATGATGATGATGATGATGATGATGATGATG 539
|||||
Qy 3505 GGGGCTGTGGCTAGTGGGAGGGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTG 3563
|||||
Db 540 GGGGCTGTGGCTAGTGGGAGGGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTG 599
|||||
Qy 3564 GGGTGGGCTCCATTTCCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3623
|||||
Db 600 GGGTGGGCTCCATTTCCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 658
|||||
Qy 3624 AAGAATATGACAGAGTGGAGAAACAGAACTTTTCTTAAGAGCTGGTAAAT 3683
|||||
Db 659 AAGAATATGACAGAGTGGAGAAACAGAACTTTTCTTAAGAGCTGGTAAAT 718
|||||
Qy 3684 CAATCTGATAC 3695
|||||
Db 719 TCATCTGATAC 730
|||||
RESULT 9
BM973838/c 644 bp mRNA linear EST 21-MAR-2002
LOCUS UI-CE-ECL1-ach-h-08-0-UI-s1 UI-CE-ECL1 Homo sapiens cDNA clone
DEFINITION UI-CE-ECL1-ach-h-08-0-UI 3', mRNA sequence.
ACCESSION BM973838
VERSION BM973838.1 GI:19591429
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 644)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: McCray, PB
McCrack Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

```

```

FEATURES
source
Location/Qualifiers
1..644
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-CE-ECL1-ach-h-08-0-UI"
/clone_1ib="UI-CE-ECL1"
/tissue_type="Lung"
/dev_stage="Adult and Fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: Lung; Vector: pT773-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-CE-ECL1 is a normalized cDNA library containing the
following tissue(s): Normal lung from adult and from fetal
day 64, day 87, week 19 and week 42. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT773-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(ori)18 tail. The sequence tag for this library is
AAGTCTTAC.
TAG_LIB=UI-CE-ECL1
TAG_TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371
and 380-383
TAG_SEQ=AAGTCTTAC"
BASE COUNT 155 a 161 c 122 g 205 t 1 others
ORIGIN
Query Match 15.7%; Score 634; DB 14; Length 644;
Best Local Similarity 99.8%; Pred. No. 3,8e-112;
Matches 634; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 3406 AGATGAGAACCAAGCCCTACAACTGCAAGCCAGCCAGGAGGGGCTGTCATCA 3465
|||||
Db 636 AGATGAGAACCAAGCCCTACAACTGCAAGCCAGCCAGGAGGGGCTGTCATCA 577
|||||
Qy 3466 CATTGAGCATGCGAATGAGCTGGGCTGGTGGTGGGCTGGTGGGCTGGTGGGAG 3525
|||||
Db 576 CATTGAGCATGCGAATGAGCTGGGCTGGTGGTGGGCTGGTGGGCTGGTGGGAG 517
|||||
Qy 3526 GGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 3585
|||||
Db 516 GGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 457
|||||
Qy 3586 CTGCTGCTCCCTCTGAGGTCTGATTTGGGATGGGCAAGAAATGACAGATGAGA 3645
|||||
Db 456 CTGCTGCTCCCTCTGAGGTCTGATTTGGGATGGGCAAGAAATGACAGATGAGA 397
|||||
Qy 3646 AACACAGAAACTTTTCTTAAGACCTGCTTAATCAATCTGATACGCTTACAA 3705
|||||
Db 396 AACACAGAAACTTTTCTTAAGACCTGCTTAATCAATCTGATACGCTTACAA 337
|||||
Qy 3706 TACAATAGTATGACAGTAAATAATATGATGCTTATATATATATATATATATCT 3765
|||||
Db 336 TACAATAGTATGACAGTAAATAATATGATGCTTATATATATATATATATCT 277
|||||
Qy 3766 TCAGGTAAAAAGCCAGACAGAAATGTATAGCCACTTCCATTTGTTCGA 3825
|||||
Db 276 TCAGGTAAAAAGCCAGACAGAAATGTATAGCCACTTCCATTTGTTCGA 217
|||||
Qy 3826 AAGAGTAGAATATTAACATATATGCTTATGATGCTTATGATGATGATGATGATG 3885
|||||
Db 216 AAGAGTAGAATATTAACATATATGCTTATGATGCTTATGATGATGATGATGATG 157
|||||
Qy 3886 ACTGATTACTTTTGGAGGGAGACAGTGGCTTGGAGCAGAGAGGAGGCTTAAAC 3945
|||||

```

Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).

Seq primer: M13 FORWARD
POLYA=Yes.

Db	156	ACTGTTCTCTTTTGGAGGGGACACAGTAGTTGAGCAGAGAGAGGAAAGGCTTTAAC	97
Qy	3946	ACTTACACCCCTTTTGTACATTTTGAAATTTTAAACATGTGACTGATTTNCCATTCAAAA	4005
Db	96	ACTTACACCCCTTTTGTACATTTTGAAATTTTGAACATGTGACTGATTTNCCATTCAAAA	37
Qy	4006	TAAACATTAATTTGGGCCCAAAAAAAAAAAAAAAAAAAAAA	4040
Db	36	TAAACATTAATTTGGGCCCAAAAAAAAAAAAAAAAAAAAAA	2

```

Plate: L1CM2521  row: 1  column: 19
High quality sequence stop: 542.
Location/Qualifiers
1..985
FEATURES
    source
```

BASE COUNT	211 a	280 c	290 g	204 t
ORIGIN				

Query Match	15.6%	Score 631.4	DB 14	Length 985
Best Local Similarity	96.2%	Pred. 1.1e-111		
Matches 734	Conservative 0	Mismatches 16	Indels 13	Gaps 8
OY 1457	GAACAGAAACGAGGCGCTCCCTCGGCGACACCACTCAGATCTTACTCGCAGTACTTT			1516
Db				
Db 14	GAGAGAGAACCGGGCGCTCCCTCGGCGGACACCACTCAGATCTTACTCGGACTACTTT			73
OY 1517	GGGGGCTTTGGGAAAGGCTGCTGCTCAGATCTATGTCCACTTGTCTCAACTATGAC			1576
Db				
Db 74	GGGGGCTTTGGGAAAGGCTGCTGCTCAGATCTATGTCCACTTGTCTCAACTATGAC			133
OY 1577	TATGTGTGGGATACGCTTTCACCCAGTAGGAGGCGCATAGAAATCGATTTCATCCAG			1636
Db				
Db 134	TATGTGTGGGATACGCTTTCACCCAGTAGGAGGCGCATAGAAATCGATTTCATCCAG			193
OY 1637	GGCTACATCACTCGGCACTTCTCTTTGGTCTTACTGGGAAGTACGGGAACCAAGTCA			1696

Db	194	GGCTACATCAGCTCGGCATTCCTCTTTGGTGGCTACTGGGAAGTACGGAAACCAATGTC	253
OY	1697	GAGCAACACCTGGGAGAGGTCCACACCCACAGCGCCACTTCACAGTGGATCTGGATGTA	1756
Db	254	GAGCAACACCTGGGAGAGGTCCACACCCACAGCGCCACTTCACAGTGGATCTGGATGTA	313
OY	1757	GCAGAGCTGGAAGAACTGGGTCTGGGCCGAGATATATGCTTTGTGCCCATGCTGTGCC	1816
Db	314	GCAGAGCTGGAAGAACTGGGTCTGGGCCGAGATATATGCTTTGTGCCCATGCTGTGCC	373
OY	1817	TGGAGCCTTGAGACACCACTGCAGAGGCTGCAGGTGACCCGGAACTGCTGGAGATGGAG	1876
Db	374	TGGAGCCTTGAGACACCACTGCAGAGGCTGCAGGTGACCCGGAACTGCTGGAGATGGAG	433
OY	1877	GAGCAGCGCGCTTCTGTGTGGGAAGGCCACCCCTGCCTACTTACTGTGGCCAGCAAC	1936
Db	434	GAGCAGCGCGCTTCTGTGTGGGAAGGCCACCCCTGCCTACTTACTGTGGCCAGCAAC	493
OY	1937	CACAGCAACAAGTGGGATCACCCCGGGGCTACCGCATCCAGATCAGTCTTACTGTA	1996
Db	494	CACAGCAACAAGTGGGATCACCCCGGGGCTACCGCATCCAGATCAGTCTTACTGTA	553
OY	1997	GAGCGCGTGTG-CCCCAAAACAGTCCATATGGCGAGAGGCTTCACAGTGGGAGAGTACCAGT	2055
Db	554	GAGCGCGTGGCCCCAAAACAGTCCATATGGCGAGAGGCTTCACAGTGGGAGAGTACCAGT	613
OY	2056	GAGCTGTGACCCAGCGG- AAGGAGAGAGAG- CCGAGTACAGCAGCG- TTTTCAATCAGA	2112
Db	614	GAGCTGTGACCCAGCGAAGAGAGAGAGAGCCCGAGTACAGCAGCGTTTTCAATCAAAA	673
OY	2113	TGACCCCTTGGGCCCCAC- - - TGTGATTTCACTGACTT- CATCAACAATGAGACATT	2167
Db	674	TGACCCCTTGGGCCCCCGCCACTGTGGGATTTTCAGTGAATTCATCAACAATGAGACATT	733
OY	2168	G- - CTGGAAGAATTTGG- - TGGCTGGGTGACACACTGGTTTT	2206
Db	734	GCTTGGAAGAATTTGGGTGGGTGGGGGAGACACTGGTTTT	776

RESULT	11
LOCUS	B0720486
DEFINITION	B0720486 966 bp mRNA linear EST 16-JUL-2002
ACCESSION	AGNCOCURF_8305121 Lupskl_symphathetic_trunk Homo sapiens cdna clone
VERSION	IMAGE:6193046 5', mRNA sequence.
KEYWORDS	B0720486 B0720486 GI:21859383
SOURCE	EST.
ORGANISM	human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 966) NIH-MGC http://mgc.ncl.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC)
AUTHORS	Unpublished (1999)
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov
COMMENT	Tissue Procurement: Dr. James R. Lupskl CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.llnl.gov Plate: LMAM3596 Row: A Column: 15 High quality sequence stop: 434.


```

/sex="male"
/issue_type="sympathetic trunk"
/dev_stage="adult, 16 yr"
/lab_host="DH10B"
/Note-Vector: PCMV-SPORT6 (Life Technologies); site_1:
NotI; site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCAGCGGTCG-3' and
5'-GACTAGTCTGATGCGAGCGGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.9 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine); available through Life
Technologies.
BASE COUNT      153 a      371 c      226 g      216 t
ORIGIN
Query Match      15.4%; Score 622.4; DB 14; Length 966;
Best Local Similarity 94.7%; Pred. No. 6.3e-110;
Matches 655; Conservative 0; Mismatches 36; Indels 1; Gaps 1;
QY 54 TCCGGAGACCCCGCCAGAGCCAGAGAGCCCGCTGCTGCGCTGAGAA 113
DB 1 TCCGGAGACCCCGCCAGAGCCAGAGAGCCCGCTGCTGCGCTGAGAA 60
QY 114 TACATTGCTCTCTTGGTGAATACGCTGCTCTCTTCTGCGGAAATGAACAGAGA 173
DB 61 TACATTGCTCTCTTGGTGAATACGCTGCTCTCTTCTGCGGAAATGAACAGAGA 120
QY 174 CAATCTGCTGCTCTCTCTCTGCGCGTCATACCATCTTCTGCTGCTGCTGCTGCTG 233
DB 121 CAATCTGCTGCTCTCTCTCTCTGCGCGTCATACCATCTTCTGCTGCTGCTGCTGCTG 180
QY 234 TGTGTGGAGAGGGGTGAGATGGGGGTGAACCCAGCAGCTTCCCATTTGCCCTGTAT 293
DB 181 TGTGTGGAGAGGGGTGAGATGGGGGTGAACCCAGCAGCTTCCCATTTGCCCTGTAT 240
QY 294 CTCCAGTGTGCGGAGCTTGGAGACACCTGCGGAGAGCCAGCTGTTTGGACACTGAGCC 353
DB 241 CTCCAGTGTGCGGAGCTTGGAGACACCTGCGGAGAGCCAGCTGTTTGGACACTGAGCC 300
QY 354 GAGAGAGGTGAGCGGTGATGCGCTTCTGACCCAGCGGCTGGGGCCAGGGCTGAGTG 413
DB 301 GAGAGAGGTGAGCGGTGATGCGCTTCTGACCCAGCGGCTGGGGCCAGGGCTGAGTG 360
QY 414 ATGACGCCAGAGCCCGGCTGCGAGCACTGCTTCTCTCACTGAGTGTGAGTGTGCTC 473
DB 361 ATGACGCCAGAGCCCGGCTGCGAGCACTGCTTCTCTCACTGAGTGTGAGTGTGCTC 420
QY 474 CCAAGGCTGACAGCCCTGCTCTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 533
DB 421 CCAAGGCTGACAGCCCTGCTCTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 534 TGGGCAATGCTCTTCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 593
DB 481 TGGGCAATGCTCTTCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
QY 594 CAGTCTGACAGCCCTGCTCTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 652
DB 541 CAGTCTGACAGCCCTGCTCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY 653 CCGTATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 712
DB 601 CCGTATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 713 AACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 744
DB 661 CACCGACACTGCGCCAGAGGTCTCTCGGGGGTTT 692

```

```

DEFINITION      602696217F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4828300 5',
mRNA sequence.
ACCESSION      BG718238
VERSION        BG718238.1 GI:13997425
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 694)
NIH-MGC http://mgs.nci.nih.gov/.
NATIONAL INSTITUTES OF HEALTH, Mammalian Gene Collection (MGC)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shireki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LLM10746 row: a column: 05
High quality sequence stop: 692.
Location/Qualifiers
1..694
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4828300"
/clone_1="NIH_MGC_97"
/lab_host="DH10B"
/Note="Organ: testis; Vector: BluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTCTTTTCTTTT-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      124 a      230 c      201 g      139 t
ORIGIN
Query Match      15.1%; Score 608.6; DB 12; Length 694;
Best Local Similarity 97.3%; Pred. No. 3e-107;
Matches 672; Conservative 0; Mismatches 14; Indels 5; Gaps 5;
QY 30 GACTACCGGGAACCTGACGAGAGTCCGGAGAGCCCGCCAGAGAGAGAGAGAGAGAGAG 89
DB 6 GACTACCGGGAACCTGACGAGAGTCCGGAGAGCCCGCCAGAGAGAGAGAGAGAGAGAG 65
QY 90 GCCCGGCTGCTGCGGAGATACATGCTCTCTTGGTGAATGAGTGTGCTC 149
DB 66 GCCCGGCTGCTGCGGAGATACATGCTCTCTTGGTGAATGAGTGTGCTC 125
QY 150 TTGCTGGAGAAATGAACGAGAGACAACTCTGCTCTCTCATCTTGGCGGTATACCA 209
DB 126 TTGCTGGAGAAATGAACGAGAGACAACTCTGCTCTCTCATCTTGGCGGTATACCA 185
QY 210 TCTTTCCCTTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 269
DB 186 TCTTTCCCTTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 245
QY 270 AGCTTCCCATGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 329
DB 246 AGCTTCCCATGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 304
QY 330 GCCAGCTGTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 389
DB 305 GCCAGCTGTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 364
QY 390 AGCGGCTGGGGCCAGGGGCTGTGTGATGACAGCCAGGCCCGGCGCTTGAGACACTGTCT 449

```


Db 453 AGCG-TTTCATCAGAAATGACCTTGGCCCCCACTGTGATTCAGTACTCATCAAC 511
QY 2156 AATGAGACATTTGCTGG-AAAGATTTGGTGCCCTGGGTGACAGCTGGTTTCTGCATAT 2214
Db 512 AATGAGACATTTGCTGGCAAAGATTCGGTGCCCTGGGTGACAGCTGGTTTCTGCATAT 571
QY 2215 CCCACATGCAGAGACATTCCTAACACAGTACTGTGGGAAAGCGCGTGGGCTTTCCT 2274
Db 572 CCCACATGCAGAGACATTCCTAACACAGTACTGTGGGAAAGCGCGTGGGCTTTCCT 631
QY 2275 CCGACCTTAACTTCTTTGAGC-AAGACCCCTCTCTACTCTGCGACTGCATCTACT 2333
Db 632 CCGACCTTAACTTCTTTGAGCAAGAACCCCTCTCTACTCTGCGACTGCATCTACT 691
QY 2334 TCCGAGGGGACAGAGTCTGGGGCTGCGAGTCAACCCCTAGCTTGCCCTGCCCAAG 2393
Db 692 TCCGAGGGGACAGAGTCTGGGGCTGCGAGTCAACCCCTAGCTTGCCCTGCCCAAG 751
QY 2394 CTGCTGCTGTGCCCGGACCTCCCTGCTTTCGCCA 2430
Db 752 CTGCTGCTGTGCCCGGACCTCCCTGCTTTCGCCA 788

Search completed: May 20, 2003, 07:58:03
Job time : 4970 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 07:58:08 ; Search time 64.4203 seconds

(without alignments)
1578.232 Million cell updates/sec

Title: US-10-081-408-2

Perfect score: 4095
Sequence: 1 MNCKTILVLLILAVITIFNL.....QAAACAPDLPAPFSGHGSN 763

Scoring table: BIOSUM62

Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4095	100.0	763	AAV03219	Amino acid sequence
2	2536.5	61.9	729	AAU84261	Human endometrial
3	457.5	11.2	248	ABP41516	Human ovarian anti
4	330	8.1	712	ABB91435	Herbicidally activ
5	327	8.0	687	ABB92504	Herbicidally activ
6	302.5	7.4	759	ABB92082	Herbicidally activ
7	246.5	6.0	300	ABB92910	Herbicidally activ
8	214.5	5.2	670	AAK94370	A. niger strain M-
9	127	3.1	863	AAK42952	Human ORFX ORF2716
10	125.5	3.1	460	ABB92909	Herbicidally activ

ALIGNMENTS

AAV03219	AAV03219 standard; Protein: 763 AA.	11	115	2.8	2675	21	AA807564	Protein encoded by
AC	AAV03219;	12	109	2.7	435	14	AA832989	Rat choline kinase
XX	21-JUN-1999 (first entry)	13	108	2.6	652	23	ABB92424	Herbicidally activ
DT		14	106.5	2.6	446	22	ABG13153	Novel human diagen
XX		15	106	2.6	617	23	AAE23407	Lysine oxidase pro
DE	Amino acid sequence of the vascular adhesion protein-1.	16	106	2.6	825	22	ABB67566	Drosophila melanog
XX		17	104.5	2.6	594	22	ABG07023	Novel human diagen
KW	Human; vascular adhesion protein-1; VAP-1; endothelial cell;	18	104.5	2.6	736	23	ABB57339	Mouse ischemic co
XX	Lymphocyte; inhibition; amine oxidase.	19	104	2.5	1250	22	AA655914	Amino acid sequenc
OS	Homo sapiens.	20	103	2.5	2224	17	AAW04254	Human factor V. H
FT	Key	21	103	2.5	2224	20	AAV49564	Human lipoprotein
FT	Misc-difference 43	22	102.5	2.5	541	20	AAV01648	Cytochrome P450 en
FT	Misc-difference 47	23	102.5	2.5	541	20	AAV01647	Cytochrome P450 de
FT	Misc-difference 679	24	102	2.5	1194	22	AA692078	C glutamicum prote
FT	Misc-difference /note-	25	101.5	2.5	587	22	ABB52856	Escherichia coli p
FT	Misc-difference 137	26	101	2.5	617	23	AAE23403	Lysine oxidase pro
FT	Misc-difference /note-	27	100.5	2.5	718	22	AAU04939	Propionibacterium
FT	Misc-difference 232	28	100.5	2.5	1253	23	AAU10648	Human L1CAM protei
FT	Misc-difference /note-	29	100.5	2.5	1257	20	AAW74152	Human L1 cell adhe
FT	Misc-difference 294	30	100.5	2.5	1544	23	AAU97541	Human phospholipos
FT	Misc-difference /note-	31	100	2.4	617	22	AAU02204	Trichoderma harzia
FT	Misc-difference 592	32	98.5	2.4	844	22	ABB53083	Escherichia coli p
FT	Misc-difference /note-	33	98.5	2.4	933	20	AAV63422	Human polypeptide,
FT	Misc-difference 618	34	98.5	2.4	933	20	AAV63449	Human secreted pro
FT	Misc-difference /note-	35	98.5	2.4	933	21	AAV73400	Human secreted pro
FT	Misc-difference 618	36	98.5	2.4	1152	21	AAV94920	Human secreted pro
FT	Misc-difference /note-	37	98.5	2.4	1494	23	AAU78460	Mouse beta-catenin
FT	Misc-difference 666	38	98	2.4	810	22	ABG04698	Novel human diagen
FT	Misc-difference /note-	39	98	2.4	929	20	AAV22189	Trimeric murine C3
FT	Misc-difference 666	40	97.5	2.4	308	20	AAV10811	Human secreted pro
FT	Misc-difference /note-	41	97.5	2.4	903	21	AAK42926	Human ORFX ORF2690
FT	Misc-difference 666	42	97	2.4	545	23	ABB91611	Herbicidally activ
FT	Misc-difference /note-	43	96.5	2.4	763	22	AAU01222	Propionibacterium
FT	Misc-difference 666	44	96.5	2.4	816	23	ABB9783	Human polypeptide
FT	Misc-difference /note-	45	96.5	2.4	1137	8	AAV71182	Sequence of herpes

FT	Peptide	/note= "potential N-glycosylation site"
FT		74...78
FT	Peptide	/note= "v8 peptide"
FT		170...174
FT	Peptide	/note= "v8 peptide"
FT		259...262
FT	Peptide	/note= "v8 peptide"
FT		264...277
FT	Peptide	/note= "v8 peptide"
FT		323...329
FT	Peptide	/note= "v8 peptide"
FT		359...361
FT	Peptide	/note= "v8 peptide"
FT		368...375
FT	Peptide	/note= "v8 peptide"
FT		384...393
FT	Peptide	/note= "v8 peptide"
FT		562...566
FT	Peptide	/note= "v8 peptide"
FT		597...601
FT	Peptide	/note= "v8 peptide"
FT		624...637
FT	Peptide	/note= "v8 peptide"
FT		1..25
FT	Domain	/note= "N-terminal tryptic peptide"
FT		5...27
FT	Domain	/note= "transmembrane domain"
XX		
PN	MO9853049-A1.	
XX		
PD	26-NOV-1998.	
XX		
PX	22-MAY-1998;	98WO-FI00429.
XX		
PR	23-MAY-1997;	97US-0862433.
XX		
PA	(BIOT-) BIOTIE THERAPIES LTD.	
XX		
PI	Bono P, Jalkanen S, Salmi M, Smith DJ;	
DR	WPI; 1999-131690/71.	
XX	N-PDSB; AAX28632.	
PT	Nucleic acid encoding a novel human endothelial cell adhesion protein designated VAP-1 - having an adhesive function and an amine oxidase function useful for manipulating VAP-1 mediated binding of endothelial cells to lymphocytes	
PS	Claim 1; Fig 1; 66pp; English.	
CC	This is the nucleotide sequence encoding the human vascular adhesion protein-1 (VAP-1) used in the method of the invention. The method involves manipulating VAP-1 mediated binding of endothelial cells to lymphocytes which comprises inhibiting the enzymatic activity of amine oxidase in endothelial cells, and potentiating the enzymatic activity of endothelial cells.	
SQ	Sequence	763 AA;
Query Match	100.0%; Score 4095; DB 20; Length 763;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 763; Conservative	0; Mismatches	0; Indels
		Gaps
OY	1 NMOKIIIVLLIAVTITFALVCVLLVGRGGDGEPSQLPHCPSPSPSAQPMTHPGOSOLF	60
DB	1 NMOKIIIVLLIAVTITFALVCVLLVGRGGDGEPSQLPHCPSPSPSAQPMTHPGOSOLF	60
OY	61 ADLSREELTAVVRFLTORLGPGLVDAAQARPSDNCFVSVELDLPKKAALAHLDGSGPP	120
DB	61 ADLSREELTAVVRFLTORLGPGLVDAAQARPSDNCFVSVELDLPKKAALAHLDGSGPP	120
OY	121 ABFAAIYFFGQPOPNNSELVVGPLPHEPSYRDVTVERRHGPLPYHRRPVLFQEYLDD	180

Dd	121	AREALATVFFGRQOPNVSELVGPLRHPSYMRDVTVERHGRLPYRRRPVLEFOEYLDID	180
Qy	181	OMIFNRRLPQASGGLHHCCFYKHKGRMLVMTTPAPRGLQSGDRATWGLYNTSGAGFLL	240
Dd	181	OMIFNRRLPQASGGLHHCCFYKHKGRMLVMTTPAPRGLQSGDRATWGLYNTSGAGFLL	240
Qy	241	HHVGLLELVNKKALDPARMTIQKVFYQGRYDLSLAQLEAOFEAGLVVWVLIIPNGTGSW	300
Dd	241	HHVGLLELVNKKALDPARMTIQKVFYQGRYDLSLAQLEAOFEAGLVVWVLIIPNGTGSW	300
Qy	301	SLKSVPPGPAPRLQFTYQGRFSVQGSRAVSSSLMTFSFGLGAFSGRPIDVRFQGERLY	360
Dd	301	SLKSVPPGPAPRLQFTYQGRFSVQGSRAVSSSLMTFSFGLGAFSGRPIDVRFQGERLY	360
Qy	361	YEISLQELATYGGNSPAAMTTRVDDGFGKXTTFLTFGVCPCRYLATVYDWHFLLESQ	420
Dd	361	YEISLQELATYGGNSPAAMTTRVDDGFGKXTTFLTFGVCPCRYLATVYDWHFLLESQ	420
Qy	421	APKTRIDAFVCFVEQNOGLPLRRHSDLYSHYFGGLAETVLVVSMSLTLNYDYVMDTVFH	480
Dd	421	APKTRIDAFVCFVEQNOGLPLRRHSDLYSHYFGGLAETVLVVSMSLTLNYDYVMDTVFH	480
Qy	481	PSGAIEIRFYATGYISAFLEFGATGKGNQVSEHTLCTVHTHSAHFVVDLDVAGLEMTW	540
Dd	481	PSGAIEIRFYATGYISAFLEFGATGKGNQVSEHTLCTVHTHSAHFVVDLDVAGLEMTW	540
Qy	541	AEDMYFVMAVPMSPBEHOLQRTQVTRKLEMEQAAFLVGSAPFRYLYLASNSNKKMGHP	600
Dd	541	AEDMYFVMAVPMSPBEHOLQRTQVTRKLEMEQAAFLVGSAPFRYLYLASNSNKKMGHP	600
Qy	601	RGYRIQMLSFAGEPLPONSSNARGFSWERYQLAVTQKKEEPSSSSVFQNDMPAPTVDF	660
Dd	601	RGYRIQMLSFAGEPLPONSSNARGFSWERYQLAVTQKKEEPSSSSVFQNDMPAPTVDF	660
Qy	661	SDFINNETIAGKDLVAVYTAGFLHIIPAEDIPNTVTVGNGVGFLLRPNFFEDDPSTYSA	720
Dd	661	SDFINNETIAGKDLVAVYTAGFLHIIPAEDIPNTVTVGNGVGFLLRPNFFEDDPSTYSA	720
Qy	721	DSIYFRGDODGACEVNPPLACLPAAACAPDLPAPFSHGFSHN	763
Dd	721	DSIYFRGDODGACEVNPPLACLPAAACAPDLPAPFSHGFSHN	763
RESULT 2			
AC	AA084261	standard; Protein: 729 AA.	
XX	AA084261;		
DT	08-MAY-2002	(first entry)	
DE	Human endometrial cancer related protein, AOC2.		
KW	Human; endometrial cancer; differential expression;		
OS	Homo sapiens.		
PN	WO200209573-A2.		
PD	07-FEB-2002.		
PF	31-JUL-2001; 2001MO-US24104.		
PR	31-JUL-2000; 2000US-221735P.		
PA	(BGHM) BRIGHAM & WOMENS HOSPITAL INC.		
PI	Mutler GL.		
DR	WPI: 2002-179967/23.		
XX	N-PSDB; ABK35481.		

PT	Diagnosing endometrial cancer comprises determining expression of	627
PT	nucleic acid molecules or expression products that are differentially	
PT	expressed in normal and malignant endometrium -	
XX	Claim 33; Page 131-134; 233pp; English.	
XX		
CC	The invention relates to diagnosing endometrial cancer in a subject	
CC	suspected of having endometrial cancer comprising determining the	
CC	expression of a set of nucleic acid molecules or expression products in	
CC	an endometrial sample suspected of being cancerous, where the set of	
CC	nucleic acid molecules comprises at least 2 nucleic acid molecules	
CC	selected from 50 fully defined sequences as given in the specification.	
CC	The nucleic acids are used as an array of at least 2 of the 50	
CC	nucleic acids bound to a solid substrate. Also included is a solid-phase	
CC	protein microarray comprising at least 2 antibodies or its antigen	
CC	binding fragments, that specifically bind at least 2 different	
CC	polypeptides from the 50 fully defined sequences as given in the	
CC	specification, fixed to a solid substrate. The methods and arrays are	
CC	useful for the diagnosis of endometrial cancer, selecting and monitoring	
CC	treatment regimes and identification of lead compounds useful for the	
CC	treatment of endometrial cancer. The present sequence is one of 50	
CC	proteins differentially expressed between cancerous and non-cancerous	
CC	samples.	
CC		
XX		
SO	Sequence 729 AA;	
QY	Query Match 61.9%; Score 2536.5; DB 23; Length 729;	
	Best Local Similarity 63.4%; Pred. No. 6.3e-247;	
	Matches 483; Conservative 90; Mismatches 154; Indels 35; Gaps	4.
QY	1 MNKTIIVLLILAVITIFALYCVLLVGRGDGDESQLPHCPSPVSAQPMTHPQSQDLF 60	
Db	1 MHLKIVLFLALSLITIFALAYVLLTSPGGS----SOPPHCPSPVSHRAMPHPGSQLDF 56	
QY	61 ADLSHEELTAWRELTORLGLVYAAAPRSDNCVFSVELDLPKKAALAHLDGSPRP 120	
Db	57 ADLSHEELTAWRELTORLGLVYAAAPRSDNCVFSVELDLPKKAALAHLDGSPRP 116	
QY	121 AREALATVEFGKOPRPNVSELVVGRLPHPSYKRDVTVERHGRLPHYHRRVLFQEYLDID 180	
Db	117 AREALATVLFGGQRPNPVSELVVGRLPHPSYKRDVTVERHGRLPHYHRRVLFQEYLDID 176	
QY	181 OMIFRELPDASGLLHNCFFYKRRGRNVLTMTARGLDSCGRATWFGLYUNISGFGFL 240	
Db	177 RHLKVELLPKARIFLSST--FMYNSTLAAVATPRGELSGRSRRTTWIGLYHNISGFLF 234	
QY	241 HHVGELELVNHNALDPARWTIOKFVQGYRYSLAKOLEKQFAGLVANVLLIPDNGTGSWM 300	
Db	235 HHVGELELLDNHADLPAHNTVOOVYTLGHYVADLQGLEHFGSGRLVYRVLPRLPNNGAS 294	
QY	301 SLKSPVPRGAPRLQFYDQGRPFVYQGSNVASSLMTFSGLCAFSGRIEDVVRQGERLV 360	
Db	295 SLKSRNSGGRPLRPLQFSPQSGQSYVOGNLVVSSLSMTFSGHGVFSGLRTFEDVRFQGERIA 354	
QY	361 YETISLOEALATIGXSSPAMTTRYVDGSGMKYTPRLRGVDCPRATVLYVVMHPTLEEO 420	
Db	355 YEVSVQECVSTYGASPKMTLYRTLDSSFGELGNSGVLRGVDCPRQATVMDVHLLVGGK 414	
QY	421 AKRTIRDAFCVVEONQGLRLRRHNSDLYSHYFEGGLAEYLVLVVRSKSTLINTYDVMDTVH 480	
Db	415 AVQLLRGAVCVVEEAGQGLRLRRHNNYLDHNFYFGGLASSALVLRVSVSGVQDYDTMDFVL 474	
QY	481 PGGALEIRPATGYSASALFGATG--KKGNOVSEITLCTVTHTSHAFKVDLDVAGLENN 538	
Db	475 PGGALEGRNHAAGYINTATLKGEBGLRLGNRGEVRLDTVTHNHFKLIDVAVGLKWN 534	
QY	539 VAAEDVEFVMAVPMSPEROLQRYTRKYLEMEBQDAVLVSGATPRYLYLASNHSNKGK 598	
Db	535 VAAEDVEFVMAVPMREHMLQRPOLTRQVLDGKEDLTATSLGSLPRYLYLASNQTNNAG 594	
QY	599 HPRGRIQMLSPAGEPLPONSSHARGFSMERIOLAVTORKEEPPSSSVFNQNDPAATV 658	
Db	595 HORG-----YQLVVTOREESQSSSIYHNDIMTPT 627	

[illegible]

CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents a human ovarian antigen of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 248 AA;

Query Match 11.2%; Score 457.5; DB 23; Length 248;
Best Local Similarity 41.9%; Pred. No. 2.7e-37;
Matches 83; Conservative 45; Mismatches 67; Indels 3; Gaps 1;

QY 305 PVPSPAPPLQFVPGQPSVSGSRVASSLWTFSGFAGSPRIEDVAFQGERLYEIS 364
DB 16 PSHVSGRLVQPHGPRFRLREGNAVLYGWSFAPRLRSSGLQVLAHFGGERLYEIS 75
QY 365 IQEALAIYGGNSPAAWTRYYVDGFGMGKYYTPPLRGVDCPYLATYVDMHFLLESQAPRT 424
DB 76 VOEAVALYGGHTRPAGMOTFYLDVGMGLSVTHELAPGIDCEPATAFLDFHYDDADDPVH 135
QY 425 IADACVFEQNGGLPLRRHSDLYS---HYEGIAETVYVBSMTLLNYDVYVMDVFEHP 481
DB 136 YRALCLTEMPGVPLRRHFSNFKGFEVYAGLKGQVLAHTSTVYNYDYIMDFEYF 195
QY 482 SGAEIRFYATGYISSAF 499
DB 196 NGVMAKKHATGYVHAXF 213

RESULT 4

ABB91435

ID ABB91435 standard; Protein; 712 AA.

XX ABB91435;

XX 31-MAY-2002 (first entry)

XX Herbicidally active polypeptide SEQ ID NO 646.

XX Herbicidal; plant; agriculture; herbicide.

XX Arabidopsis thaliana.

XX WO200210210-A2.

XX 07-FEB-2002.

XX 28-AUG-2001; 2001WO-EP09892.

XX 28-AUG-2001; 2001WO-EP09892.

XX (FARB) BAYER AG.

XX Tietjen K, Weidler M;

XX WPI: 2002-269010/31.

XX Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plant with nucleic acid or amino acid sequences from non-plant
XX organisms.

XX Claim 5; SEQ ID NO 646; 261bp + Sequence Listing; English.

XX The invention relates to identifying target proteins
XX (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX aligning and comparing nucleic acid or amino acid sequences from plant

CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.

XX Sequence 712 AA;

Query Match 8.1%; Score 330; DB 23; Length 712;
Best Local Similarity 21.7%; Pred. No. 1.3e-23;
Matches 172; Conservative 107; Mismatches 346; Indels 166; Gaps 28;

QY 7 LVLLIAVITITFALCYLVYRGDGG-----EPGSLPHCPSPSPAQWTHPG 55
DB 9 LFLFFSFLILFATYSWVF--GDSGFLEGRVVKRTIGSNQVH--VDHSLERPHHPL 62
QY 56 OSQPLFADSLREELTAVMRFLTORLPGVDAQAARPSDNCVSEVETOLPPKAAALAHDR 115
DB 63 DLUTREIR-----VRIILSNHDPGFGSSAT-----ISMALDEEK-SRVQWKK 109
QY 116 GSPPARREALVFFGROPQPVSELVGPPLPHPSYMDVYERHGGPLPYRRPVYFQE 175
DB 110 GKLLSRRAAVAYVGGQTHETVLDLDSGRV-----VSDVINRTSGYPI-LTANDVFAAS 163
QY 176 YLDIDQMIIFRELPOASGLLHCCFKHGRMLVMTTAPRGLSGDRAT-----WGL 229
DB 164 QVPLKSLFENRSI-----EARGVKSDSLACTIPFGWGS- 197
QY 230 YNNSGAGFFLHHVGLLELVNHRKALDPAKWTIQKYQGRYDLSLAQENF---EAGLY 286
DB 198 -----SDEGRVY-----RVQCFITQGTNYFPRPLEGLVTVYDIDL 236
QY 287 NVLLIPDNGTGSWSLKSVPVPPAPPLQF-----YQGPRESVO-G 327
DB 237 EVIKIIDKG-----PIPIKASGTEYRFYQVKNRPHMDRINPISMEQPDPSFVEDG 289
QY 328 SRVASSLWTFSGFAGSPRIEDVAFQGERLYEISIQEALAIYGGNSPAAWTRYYVD 386
DB 290 HLVKMANWVYHVKADQORAGMIT-----SOATVRDSETEGEPSPVMTKGFSELFVYMDP 343
QY 387 -----GFGMGKYYTPPLRGVDCPYLATYVDMHFLLESQAPRTIRDACVFEHP 434
DB 344 EGGWYKGYMDAGELGLPTAMPVPLVNDCPNRSYIDGVFASPGOKRIVQNMICLFER 403
QY 435 NOGLPLRRHSDLYSHYFGLAET---VLVYRSMSTLLNTYWDYVHFGAIEIRFY 490
DB 404 YAGDISWRHSEILFPAN--ADIRESRPKYTLVARMATISGVNYIYIDFWEQDTGLIRYVA 461
QY 491 ATG-----YISSAFLEGATGKGNVSEHTLGTVHTSHAFKVDLTVAGLENNWAE 542
DB 462 ASGMLAVKCTPYDNDVDDLDDREDDAGPLISENVIGVYHDFITFHMDIDGPMN---N 517
QY 543 DMVFPVM---AVPWSPEHQRLQVYTRKLLMEDEQAALVGSATPRYLYLAS--NHSNKG 598
DB 518 SLVKVHLEKQRYVPTGKSPKSYLKVKYIAKTEKDAQIKLSYDPEFHIIVNKRSRVG 577
QY 599 HPRGRTIQMLSTFAGEPLPQNSSMARGFSEMRQQLAVTQKKEEPPSSSVFNQNDPAPYV 658
DB 578 NPAGRYIVPGNAASILDHDDDEPQIRGAFTNNQIOWTTPYRNSQYAGGLIYQSGDDTL 637
QY 659 D-FSDFINNETIAGDLVAMVTAAGFLIHPADIPTVTVAGNGVGFLLPYNFFEDPSF 717
DB 638 QVMSD--RRSTENKDIYMTLGFHHYVQEDYPMPIV--AASLELKPAFFESNPIL 693
QY 718 YSADSIYFRGD 728
DB 694 GSAP--FFEKD 702

RESULT 5

ABB92504

ID ABB92504 standard; Protein; 687 AA.

XX ABB92504;
 XX
 DT 31-MAY-2002 (first entry)
 XX
 DE Herbicidally active polypeptide SEQ ID NO 1715.
 XX
 KM Herbicidal; plant; agriculture; herbicide.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200210210-A2.
 XX
 PD 07-FEB-2002.
 XX
 PF 28-AUG-2001; 2001WO-EP09892.
 XX
 PR 28-AUG-2001; 2001WO-EP09892.
 XX
 PA (FARB) BAYER AG.
 XX
 PI Tietjen K, Weidler M;
 DR WPI: 2002-269010/31.
 XX
 XX Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms -
 XX
 PS Claim 5; SEQ ID NO 1715; 261pp + Sequence Listing; English.
 XX
 XX The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.
 CC
 SQ Sequence 687 AA;
 Query Match 8.0%; Score 327; DB 23; Length 687;
 Best Local Similarity 21.6%; Pred. No. 2,4e-23;
 Matches 171; Conservative 113; Mismatches 314; Indels 194; Gaps 31;

DB 281 HIRPDQACMI-----SQATVRDSKTEGARSVMKGFASSELFPNMDGECNYSKAYM 334
 QY 386 DGG-FGKGYTPTLRGDCPYLATYVDVHFLLESQAPKTRDAFCVFEONOGLELRHH 444
 DB 335 DAGEFGLGSSMPLVLPNDPPRNAIYIDGFFASPEGPILOPNMICKFERVADTSRHS 394
 QY 445 SDLYSHYFGL-----AETLVYRSMSTLNDYMDVTFHPSGAIEIRFVATG----- 493
 DB 395 EIL-----LPGVDIRESPRAKTLVARMACSGVNDYIDFMEFQMDGVIRVTVAAAGMLVXK 450
 QY 494 ----YISSAFLEPGATGKGNQVSEHTLCTVTHSHAFKVDYDVGLEMMVAEDVVFPM 550
 DB 451 GTAENVEDLGEKEDDSGLPISENVIVVDHFIHFHLDMDIDGSAN-----NSFVKVHLE 506
 QY 551 VPMSPEHQLOR---LQVTRKLEMEFOAAFLVGSATPRVLYLAS-NHSNKGHPRGYRIQ 606
 DB 507 KQRLPGESESRKSLKAKKYAKTEKDAQIKMSLYDEYFHLVAPNLSRLGNAGYKLV 566
 QY 607 MLSFAGEPLPONSSMARGFSEWERYOLAVTQKEEPESSSVFNQNDPMA----- 655
 DB 567 PGNAASGLDHDHDPQMRGAFNNQIWYTR-----YNSQEWAGGLMYQSRGE 615
 QY 656 PTVD-FSDFINNETIACKDLYAWYTAGFLHHPHADIPTNYGNGVGFELRPVNFEDD 714
 DB 616 DTQVWSD--RDRSIEKKDILWYTLGFHHVPCQEDPVPMTIAS--SFLKPVNFESN 671
 QY 715 P-----SFYSAD 721
 DB 672 PVLGISPFPEKD 683
 RESULT 6
 ABB92082
 ID ABB92082 standard; Protein: 759 AA.
 XX
 XX ABB92082;
 AC
 XX 31-MAY-2002 (first entry)
 DT
 XX Herbicidally active polypeptide SEQ ID NO 1293.
 DE
 XX Herbicidal; plant; agriculture; herbicide.
 KM
 XX Arabidopsis thaliana.
 OS
 XX WO200210210-A2.
 XX
 PD 07-FEB-2002.
 XX
 PF 28-AUG-2001; 2001WO-EP09892.
 XX
 PR 28-AUG-2001; 2001WO-EP09892.
 XX
 PA (FARB) BAYER AG.
 XX
 PI Tietjen K, Weidler M;
 DR WPI: 2002-269010/31.
 XX
 XX Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms -
 XX
 PS Claim 5; SEQ ID NO 1293; 261pp + Sequence Listing; English.
 XX
 XX The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant

CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.

XX Sequence 759 AA:

Query Match 7.4%; Score 302.5; DB 23; Length 759;
 Best Local Similarity 22.5%; Pred. No. 8.5e-21;
 Matches 166; Conservative 98; Mismatches 327; Indels 147; Gaps 29;

```

OY 43 SVSPSAQPTWHPGQSOLFADIS-----REELTAANKRFLTQRLGPGIYDAQAAP 91
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 71 SVMRTER-KRPLDPLSAEISVAATVRAAGANPEVDGRF-----IEVASERP 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 92 SDNCVFSELD-LRPKAALAHLDGSP-----PRAEALAIYFGROPQPNSELYVG 144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 DKQVALADAFPPFPQSLPRPKSGVIMPKLPPRRARLVVY--NOKSNERTSWIYA 177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 145 PLPHRYMDYTVERRHG-----PLPHRRPVLFQELDIDIMIFNNELPOAGSL 194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 178 -----LSEVHAATRGHHGRVSSQVIRPDVQPMALAEAEALY--KDFPP----- 224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 195 LHHCCFYHRRRNLYMTTAPRGLOGSDRA--TWFGLYINISGAGFPLHHVLELLVNH 251
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 225 -----FIEAMR-----RGIEDMDLVWVDPWCVGHSSEDA----- 255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 252 KALDPARMTIKVFYQGRYDSLQLEAFEGLVNVLID-----NGTGGG 299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 256 -----PSRLKAPLYICRF--DSDSMENGIVRPVEGITVLDMOMNVIEEDRESRG- 308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 300 WLSKSPVPGAPPLQ-FYPOGPRESVQGSRVASSLMTFSGLAFSGPRIFDVRF----- 354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 309 -----VDRSDVKKLQIOPBSPFRVGVYEMQKMFRIQFTPRELVIHSVAIVGGS 362
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 355 QGER-LYVEISLOELALYGG-NSPAAMTTRYVDGFGMGKTTPLNGVDCPYLATYVD 412
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 363 RGRPPVAHRLSFEVENVYVYGDPPNEPHYKNAFDAGEDGLGNASHLKGCCDLGSIKYFD 422
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 413 WHFLLESAQPTIRDAFCVFQONOGPLRRHSHDLXSH--YFEGLAET-----VLVRSMS 466
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 423 AHFNFTGTGVTIENCYVLHEDHGI-----LMKHQDRTGLAEVRRSRRLVSELC 474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 467 TLNVLDYMDTVFHPGSAIEIRFYATGYIS-SAFLEGATGYXQNOVSEHTLGTVTHSAH 525
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 475 TVANVEYGFYHMFQODGKIEAEVKLTGLISGALPGFETRYKGTITAGGLAVPAHQHFI 534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 526 FKVDLDY-----AGLENWMAEDMVFVPMAPVPSPEHQLOR--LOYTRKLLMEEOALV 579
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 535 ARMDISVDCKPAEAFNOY-----VEENVYVDERGENNHNNAFYAEKLLKSEAVAMRDC 589
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 580 GSATPRYLTYLASNHS-NKMGHPRGYRIOMLSFAGEPLPONSMMARGFEMERQOLAVTQR 638
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 590 DPLSRHIVKNTKRVNTGTGTLTKLPVGSNCPLARPEAKFLRRAFLKHNLMVNTXIA 649
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 639 EEPSSSSVFNQNDPMAPTVDFSDFINNETIAGKDLVAVMTAGFLHIAEDIPVTYVG 698
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 650 PDEFKPGGEFNPQNPRAEGELATWVKQNSLSESDVLMVYFGIHHVRLDEMP--VMPV 707
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 699 NGVGFELRPYMFPEDEPS 716
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 708 EHIGFTLMPHGFFNCSPA 725
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 7

ID ABB92910 standard; Protein: 300 AA.

XX ABB92910;

XX 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 2121.

KW Herbicidal; plant; agriculture; herbicide.

XX Arabidopsis thaliana.

OS WO200210210-A2.

XX 07-FEB-2002.

XX 28-AUG-2001; 2001WO-EP09892.

XX 28-AUG-2001; 2001WO-EP09892.

XX (FARB) BAYER AG.

XX Tietjen K, Welder M;

XX WPI; 2002-269010/31.

PT Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms -

PS Claim 5; SEQ ID NO 2121; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins

CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.

XX Sequence 300 AA;

Query Match 6.0%; Score 246.5; DB 23; Length 300;
 Best Local Similarity 25.8%; Pred. No. 8.5e-16;
 Matches 74; Conservative 53; Mismatches 133; Indels 27; Gaps 7;

OY 457 ETVLVRSMTLNVYMDTVFHPGSAIEIRFYATGYISSAFLEGATG----- 506

DB 11 KVTLVVRVAVSANGNDYIIDYEFQDGVRAKVGLSGWL--MKGTYEKNKNGYKXDK 67

OY 507 -----YGNQVSEHTLGVTHSAHFKVDLVAGLENNVMAEDMVFVPMAPVPSPEHQL 560

DB 68 GNEBELYGTILSENVIGYIHDHYTFYLDLDVGDGNSFVKNKROETAGGESPRKSY- 126

OY 561 RLOYTRKLLMEEOAALFVGSATP-RYLYLASNHSNKKMGHPRGYRIOMLSFAGEPLPONS 619

DB 127 -MKAVRNIVKTEKDGQIKLSLYDSEHYVNPCKTTRVGNFTGIXVPRATPAASILDHDD 185

OY 620 SMARGFSEWRYQAVTORKEEPPSSSVFNQNDPMAPTVDFSDFINNETIAGKDLVAVM 678

DB 186 PPOKGAFTNNQIWTVPYKSEQMAASGLFTYQSHODDTLAVMSD--RDRDIENDDIYVY 243

OY 679 TAGFLHIAHADIPNTYVVGNGVGFELRPYMFPEDEPSFTYADSITF 725

DB 244 TLGFHHIPCOEDFPIPTVSS--SFDLPVNFPERNPILKAAPNEY 288

RESULT 8

ID AAR94370 standard; Protein: 670 AA.

XX AAR94370;

XX 20-NOV-1996 (first entry)

DE A. niger strain M-62 amine oxidase.

XX Strain M-62; amine oxidase; formation; hydrogen peroxide;

KM butyl amine; benzylamine; oxidative deamination;
 KW recombinant production.
 XX Aspergillus niger.
 XX JP08070872-A.
 XX 19-MAR-1996.
 XX 05-SEP-1994; 94JP-0211526.
 XX 05-SEP-1994; 94JP-0211526.
 XX (TOKU) TOKUYAMA SODA KK.
 XX WPI: 1996-203152/21.
 DR N-PSDB; AAT14174.
 PT Gene encoding amine oxidase - used in formation of hydrogen peroxide
 PT from butyl amine, or benzylamine by oxidative deamination
 PS Claim 3; Pages 9-11; 12pp; Japanese.
 CC The present sequence is the A. niger strain M-62 amine oxidase
 CC (AO), which is used in the formation of hydrogen peroxide from
 CC butyl amine, or benzylamine by oxidative deamination. The AO gene,
 CC the sequence of which was determined by sequencing a restriction
 CC enzyme map of an AO gene contg. cDNA, can be used for the
 CC recombinant prodn. of AO.
 SQ Sequence 670 AA:
 Query Match 5.2%; Score 214.5; DB 17; Length 670;
 Best Local Similarity 22.3%; Pred. No. 5,6e-12;
 Matches 168; Conservative 86; Mismatches 301; Indels 197; Gaps 37;
 QY 53 HGGQSLPRLDLSRELTAVMRPLTORLPGVLYAQAARSDNVFVELQLPKA----- 107
 DB 4 HP-----LAILSEETNIAKNVLAQHPTVIDERE-----IYSEPKAQLLEF 48
 QY 108 AALAH---LDRGSPPREALAA---IVFGRGP-----QPNVSELVVGPLPHPS 150
 DB 49 LALEHSGRLSPSPRRPRALACQDYVIGNDRISEFESVVDVGTQRQVRRVVGKHEHAS 108
 QY 151 YMR--DVTVER-HGGPLRYHRRPLVLEQYL-DIDQMIENRELPOA-----SGL--- 194
 DB 109 LTLSEFDTLVERCFASP-----LFOKALADFD-----LPEGEVVIIEPYGIDYV 153
 QY 195 -----LHNCCEYKHKGRN-----LVTMTIAPRGOSGDRATWFGLYNNISG 235
 DB 154 EEKRRYFQGLCFATYDRKKNPDANFYSYPLPLPVMDGAPGSHGSPSR-----HCKG 208
 QY 236 AGFELHNVGELLELVNKHALDPARMTIQKVEYQGRYVDSLQLDAQEFAGLVNVLLIPDNG 295
 DB 209 EGIETDQTFKRDIIIGHCKDSP-----IYVELL 234
 QY 296 TGGSWSLKSPVPPGAPRPLQFYPOGPRFSV-QGSRVASSLWTSFGL-----GAFSGPRIF 350
 DB 235 PGCTRDLDKPLN-----VVOQEPSEFRITESLSIVEMQKMRFRVAFNPRGCVHSQTSW 287
 QY 351 DVAFQGERLVYEISLQALAIYGGNSPAAFTTRVYD-GGCGMKYTTPLRLRGVDCPLAT 409
 DB 288 ---YDRSVLYRLYSVSEMTVPYADPRPFHRKQAFDGGDGGGMANNLSIGCCCLGVIK 344
 QY 410 YVDHNFLESGAPKTIIDAFCEFEQNOGLPLRRHNSDLYHYFGGLAETV-----LVVRS 465
 DB 345 YEPAAVMTGADGSAKKMNAICLHEDONGIGK-----HSNMRKGRAVYTRHRELVVQFI 398
 QY 466 STLLNTDYVDVYFHPGSAIEIRFYATGYISSAFLGATSKY-----GNQVSE---HTLG 517
 DB 399 IITLANYEYFAVKPDGSGITGRVACHGYLE-----RCQHRCWQGOQRYVQRQMRBVG 451
 QY 518 ---IVHHSAHFKVDLDVAGLENNVMAEDMVYVMAVPSPEHQDLQVTRKLEHDEQ 574

DB 452 PEPPAHFLCAHRP---GYTPNNNSVOVESHVPPNANATVNPNGNRYKVT-----ETMR 503
 QY 575 AAFVGSATP---RYLYLASNH-----SNKMGHPRGYRIOMLSFAGEPLPONSSNARGPS 626
 DB 504 AGFF---DAAPELNRTVKVMNPKKNPISOK---PYGKFIPLATQRLRLADPNSIOARRAQ 558
 QY 627 WERYOLAVTQRKEEPPSSSVFN-QNDPAPRYVDSDF-NNETIAGKRLVAMVYTAGFLH 684
 DB 559 FAQHHVWYTKYRDGELYAGRYTLQSQEIEGV--SDAVKRGDSVYDIDVVMVSFFGLTH 616
 QY 685 IPHAEDIPRTVTVNGVGFELRPYNFDEDP 716
 DB 617 NPREVDP--VAPVELFQMLMRPADEFTANPS 646
 RESULT 9
 AAB42952
 ID AAB42952 standard; Protein; 863 AA.
 XX AAB42952;
 DT 08-FEB-2001 (first entry)
 DE Human ORFX ORF2716 polypeptide sequence SEQ ID NO:5432.
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnerability; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineumatic; antihypoid;
 KW antiviral; antibacterial; antifungal; antipneumatic; antihypoid;
 KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; nocturnal haemoglobinuria; burn; wound;
 KW thrombosis; contraceptive.
 KW Homo sapiens.
 OS Homo sapiens.
 PN WO200058473-A2.
 XX 05-OCT-2000.
 XX 31-MAR-2000; 2000WO-US08621.
 XX 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX (CURA-) CURAGEN CORP.
 PA Shinketsu RA, Leach M;
 PI WPI: 2000-602362/57.
 DR N-PSDB; AAC77161.
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX Claim 11; Page 4615-4617; 5507pp; English.
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnerability;
 CC antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;

CC antiadipetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;
 CC antihypertoid; antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC graft vs host disease, neurodegenerative disorders, osteoarthritis,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 863 AA;

Query Match 3.1%; Score 127; DB 21; Length 863;
 Best local Similarity 22.3%; Pred. No. 0.0061;
 Matches 100; Conservative 34; Mismatches 126; Indels 188; Gaps 22;

QY 30 GDGGEPSQLPHCPSPVSP--SAQPTNHP-----GOSQLFADISREELTAVMRELTQRLGPG 83
 DB 388 GAGGAPASATPTPALSPGRSLRPHILPLILRGAPAPLTDACQCMCKSLRGAGQPLGP-- 445
 QY 84 VDAQAARPSDNCVSVLEQLPKAAALAHIDRCSPARE-----ALAIYFEGQOPQPNV 138
 DB 446 -----DMESPRLPPPLSLRPGAGAPPPKPARMLALALERRAQVAEOQ 491
 QY 139 SELVGDPLPHPS--YMRDVTVERHG-----GLPYHRRPV 172
 DB 492 SQGECGGTTPPASOSPFHRSLSLEVGGERPLTSGSGPPNSLAHGCANVPGPPLPR--- 548
 QY 173 FOEYLDIDOMIFNRELPOASGLLHCCFYHNRGNLVTMTAPRGLOSGDRATWFGLYYN 232
 DB 549 -----QOSGSGSL-----RSQRPMTGSRGLRGP----- 572
 QY 233 ISGAGFLHNVGELLYNHRKALDPAWTKVFGGRYYSLAQLAEQFAGLVNVLIP 292
 DB 573 -----AOVSQRLRAG----- 582
 QY 293 DNGTGS-----WSLKSFPV-PG---PA-----PPL-----QFYPOGPRFSVQ 326
 DB 583 -----GGGRDAPRAAAGSPSCVSPSQVPPRGFFSPAPRCLPFLGVPKGLVPLGP--SFQ 637
 QY 327 GSRVASSLMTFSGLGFSGPRITDVRFQGERLYVEISLQDALIYGCN---SP-AAWTT 382
 DB 638 PSSPA-PVWRSLSL-----PPAPLD---RGENTLYEIGASEGSPYSGPTRSMSPFRSMKP 688
 QY 383 RYVDGSGMGKYYTPTLRGVDCPYLATY 410
 DB 689 DRLNASTGMGOSPPPLHRSPD--FLLSY 714

RESULT 10
 ABB92909
 ID ABB92909 standard; Protein; 460 AA.

XX ABB92909;

XX 31-MAY-2002 (first entry)

XX Herbicidally active polypeptide SEQ ID NO 2120.

XX Herbicidal; plant; agriculture; herbicide.

XX Arabidopsis thaliana.

XX WO200210210-A2.

XX 07-FEB-2002.

XX

PF 28-AUG-2001; 2001WO-EP09892.
 XX 28-AUG-2001; 2001WO-EP09892.
 PR (FARB) BAYER AG.
 PA Tietjen K, Weidler M;
 XX WPI; 2002-269010/31.
 DR
 XX
 XX
 PT Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms -
 PS Claim 5; SEQ ID NO 2120; 261pp + Sequence Listing; English.
 XX
 XX
 CC The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.

XX Sequence 460 AA;

Query Match 3.1%; Score 125.5; DB 23; Length 460;
 Best local Similarity 20.7%; Pred. No. 0.003;
 Matches 101; Conservative 49; Mismatches 186; Indels 153; Gaps 22;

QY 31 DGGEPSQLPHCPSPVS-----PSAOPWTHPGOSQL---FADLSREELTAVM 72
 DB 39 DCRKSSSPFCASNFLENKNTREPSKNIPTKPKNDHVSDFRNHPLDPLTVMEINKVR 98
 QY 73 RFLTORGLPGLYDAQAARPSDNCVSVLEQLPKAAALAHIDRCSPAREALAIYVEFGR 132
 DB 99 SILSSH-----ALFASRVPHLNSVYLEEPDK-NLYROMEKGDQLPRKASVYARVGG 150
 QY 133 QOPNVSELVVGRLPHPSYMRDVTVERHGRLPYHRRPVLFQEYLD-----IDOMTFNR 186
 DB 151 N-----SHLLIVDL---STSR---VDQADSPVPSGYPVYTSSEMDASAASFENADENR 199
 QY 187 ELPOASGLLHCCFYHNRGNLVTMTAPRGLOSGDRATWFG-----LYYNI 233
 DB 200 TI-----NSRGVNLTDVICIP--ISSG---WFGNDDNTKRYTKIOCFSTQ 240
 QY 234 SGAGFLHNV-GLELLYNHRKALDPAWTKVFGGRYYSLAQLAEQFAGLVNVLIP 292
 DB 241 DTPNFYMRPIEGTLT-----FDLDTKRILET 268
 QY 293 DNGTGSWSLKSVPYPPGAPPLQF-----YPOGPRFSYOGSRVAS-S 333
 DB 269 DTGQ-----SIPFGSTNMDYRSTLPNHDKTRPLNTISLEQRGSGFYIEDNHLKWA 322
 QY 334 LMTFSF-----GLGAFSGPRITDVRFQGER-LYVEISLQDALIYGCNSPAAWTTTYVD 386
 DB 323 NMEFHLKPDPRAGL-IMSQVKIHDPDQETREVMYKGFVSELEFPYVDPSEAWTFKTYMD 381
 QY 387 GG-FGMGKYYTPTLRGV-----DCPYLATYDWMHFILESQAPKTIROAFCYFEQ 434
 DB 382 AGEYGFGLQAMPLEPAEPKNLFTGLNDPCRNANVYMDGFAAADGTPYVRENMICVFEES 441
 QY 435 NOGLPLRRH 443
 DB 442 YAGDIAMRH 450

RESULT 11
 AAB07564
 ID AAB07564 standard; Protein; 2675 AA.


```

Query Match 2.8%; Score 115; DB 21; Length 2675;
Best Local Similarity 18.9%; Pred. No. 0.65;
Matches 172; Conservative 107; Mismatches 321; Indels 310; Gaps 42.

QY      48 AQPWTHPGOSQLFADLSREELTAANKFLYORLGRFVGLDAQAQR---PSDNCYFSEYLOLP 104
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1329 ARPGNHDPADRLVRIAREAAVTV--HFVPSMLTFLETALRGTRLRALRVVSGSELP 1387

QY      105 PKAA--ALANDR---GSPRRAREALAIYFG--ROQOPVSELVGRLPHPSIMRYUYE 158
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1388 AAANVRAGGLDARLYNLGPTFEAAVDVYAMPCRPREP-----GRVEIGLPIANTTAE 1440

QY      159 RHGG---PLPYHRRPPLFEQYLDIDQIMINRELPAQSGLLHHCSEFKHGRLVMTWTAP 215
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1441 VLDGRLRP---RPVGGELYL-----GGACLAH--GYHNDPALTAARLPAP 1483

QY      216 RG---LQSGDRATWELGYLVNIGAGFLLHHVGLLELLVNNKALDPAKMTIOKVFYQGRYUD 272
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1484 GGGRRRTGDL-----VRRAGACALVFRGRTDQVNIIGGIYRVP-----GEVAE 1527

QY      273 SLAQLEAOFEAGLVNVLIPDNGTGGSMKISVPPGP----- 310
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db	1528	ALBAL-----PGADAAVVPDHGRLAAYAVADPVGPARAPALBDALRRLLPHGLYPAAL	1582
Qy	311	-----APPLQFYPOGRPFSSVQS--RAVASLMTFSFOLGAFSG	346
Dd	1583	TLLDRPLTPAGLYLRALLPHPSAPP-----PPGGPRPTTGTERLVARW-----	1627
Qy	347	PRIEDVRFOGERLYVE-ISIGDALIYGNSPMAATTXYVDGFGMGKYTTLPLTRGVDP	405
Dd	1628	-----AERLGREVVGDNDGFSSLGSDSVRALGSVALALBAAGLPVTYDILRLPTVA	1678
Qy	406	YLATYYDMH-----FTL--ESQAPKTIIDAFVCFEONOGPLRNRHSGLYS	449
Dd	1679	ALARHADDERADRPRAROETPRPGFALCSERAAGVPLEDAVPMSPMAQRAYLEHRDNHGYE	1738
Qy	450	HFFGGLA---ETVLVBSMSTLIN-YDYV---MDTYFHPSGA-----TEIRFY	490
Dd	1739	VYVTSVAVS+PLDRTRLAAAVDLRLDRHAUYLRSSFDSLVSHEPQTOLVTHLPTLEY---	1795
Qy	491	ATGYISSAFLFG-----ATGKYGNQVSHHTLTGVHTH---SAHFKYDLDAVGLENMYMA	541
Dd	1796	----VESSDPAGFDANLHAERKRPLDVGTGPLARETTAHDACAGAQRLYVSSFALDGKCYA	1851
Qy	542	EDMWVF-----PAWP-----MSPE-----HQI	559
Dd	1852	TVLTELLROYKSALRGCAPLSLRPARASTREFVALERAAQHDPANHEFPRTTELACARPHL	1911
Qy	560	ORLOVT-----RKLEMEEOAA-----FLVG-----SATP	584
Dd	1912	PRRPVPRPPGDDGIQRHHRHVYPVEDTVAKGLSLAGELGVLKHVLLGLVHLRYVRALSQP	1971
Qy	585	RILYTLAHSNHNKKGH-----PROGRICQLSPAGEPLPQNMSGARGSWERYQ	631
Dd	1972	DVIYAVEVTHRLERLHHODRVLVGFNNILPLRORVDSGMADIARAHAHAEAARTGETRKP	2031
Qy	632	LAVTORKEEPPSSSSHYNONDPAPRYDESDFINNETIAGKDLYVMVTAAGFLHIPHAEDI	691
Dd	2032	LAQAQR---DHGAAGLF-----TLFVTHHFLYALALDDGMA---VSDLRARDQTYV	2079
Qy	692	PNT-----VTVNGVGEEFR--PYNEFDEDPSEYSADSITYRGDODAGACEVNPACLAP	743
Dd	2080	PLTAHFVWDATDGGGLRLLESDDRPREPDQOVAEFPA---YYRR-----AL	2122
Qy	744	QAACARDLP	753
Dd	2123	RAAADARHPR	2132
RESULT 12:			
AAR32999			
ID	AAR32999	standard; Protein: 435 AA.	
XX	XX		
XX	AAR32999;		
DT	14-JUN-1993	(first entry)	
DE	XX	Rat choline kinase.	
KW	CK; epitope; tumour; diagnostic; antibodies; distribution; liver;		
KX	cranial nerve disease.		
OS	Rattus rattus.		
FH	Key	Location/Qualifiers	
FT	Peptide	139..225	
FT	Peptide	278..327	
FT	Peptide	328..360	
FT	Peptide	398..412	
PN	JP05015367-A.		
PD	26-JAN-1993.		
PF	17-SEP-1991;	91JP-0262516.	

XX 14-SEP-1990; 90JP-0242528.

XX (MITK) MITSUI TOATSU CHEM INC.

XX WPI; 1993-070176/09.

DR N-PSDB; AA037687.

PT Rat choline kinase having specified DNA sequence - useful for
PT diagnosing cranial nerve diseases, liver diseases, tumours, etc.

PS Claim 1; Page 21; 26pp; Japanese.

XX The choline kinase (CK) gene may be obtd. from rat brain, human
CC brain, human liver, chicken liver, egg, soy bean, etc. The rat CK
CC gene is shown, and can be isolated by PCR using oligonucleotides
CC corresp. to partial CK sequences as primers. The oligonucleotides
CC or the amplified DNA fragments, can then be used as probes to detect
CC CK coding sequences. Rat CK may be purified by MMA-agarose affinity
CC chromatography. A large amt. of CK may be produced and antibodies
CC against CK may be obtd. for use in detecting the distribution of CK
CC in tissues or cells, and the diagnosis of cranial nerve disease,
CC liver disease and tumours.

XX Sequence 435 AA;

Query Match 2.7%; Score 109; DB 14; Length 435;
Best Local Similarity 21.5%; Pred. No. 0.13;
Matches 79; Conservative 41; Mismatches 150; Indels 98; Gaps 14;

QY 268 GRVYDLSLAQLEAFACGLVNVVLIIPDNGTGGWSLKSPPVPPAPLQFYPOGFRFSVQG 327

DB 33 GQQRDAGELESKOLGSRQPLALPPPPP-----PLPLPPPPPLADEQAPR----- 82

QY 328 SRVASSLWTSFGLGAFSGPR-----IFDVRFQGERLYVEISLOEALAIYCGNSPAATT 382

DB 83 TRRAVYLMCKEFLPGARGLREDOFHISVIRGILRNMLFCOSLPDSIASV-GDEPRKVL 141

QY 383 RYVDGFGMGKTYTPPLRGVDCPYLATYVDWHFLESQAPKTIIDAFCEFO---NOGLP 439

DB 142 RL-----YGAILMGAEMVLESVMFAILERSIGPR---LVGIFQGLREOPIP 188

QY 440 LRRHSD--LXSHYFGGLAEVLYVRSNSTLNDYVDVYFHPSGALIEIRFYATGYISS 497

DB 189 SRRIDTELCLPDISAIEAKMATFHKMKPFNKPKV----- 226

QY 498 AFLFGAGKXGNOVSEHTLGTVHTSAHEKVDLVAGLENNVMAEDVFMAYPMSPEH 557

DB 227 --LGTMEKTYLNOYLR-----LKSREARVQQLHK-----FLSYNLPLELEN 266

QY 558 QLORLQYTRK-----LLE---MEEOAALFVGSATPRYLTLA---SNHSN 595

DB 267 LRSILOYTRSPVYVCHNDCCGCGNILLLEGCGNSERKOKMLIDFEYSSTVYNGFDIGNHFC 326

QY 596 KWGHPRGY 603

DB 327 EWMYDYTY 334

RESULT 13

ABB92424 standard; Protein: 652 AA.

AC ABB92424;

DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 1635.

XX Herbicidal; plant; agriculture; herbicide.

OS Arabidopsis thaliana.

PN MO200210210-A2.

PD 07-FEB-2002.

PF 28-AUG-2001; 2001WO-EP09892.

PR 28-AUG-2001; 2001WO-EP09892.

PA (FARB) BAYER AG.

PI Tietjen K, Weidler M;

DR WPI; 2002-269010/31.

PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms.

PS Claim 5; SEQ ID NO 1635; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.

SQ Sequence 652 AA;

Query Match 2.6%; Score 108; DB 23; Length 652;
Best Local Similarity 20.1%; Pred. No. 0.32;
Matches 122; Conservative 56; Mismatches 190; Indels 240; Gaps 28;

QY 11 ILAVTIFALVCVLYVGRGDGGEPSQLPHCPVSPSAQPTHPQSOLFADLSREELTA 70

DB 151 ILVITLITLCLCKKKRRRHDEAAVYVPPRPGPKA-GGPGGQQQYMQQN----- 203

QY 71 VMRFLORLGGLVDAADARSDNCVSVLEQLPPKAAALAHDRGSPPRALAIYVF 130

DB 204 -----ASRPSDHNHYVT--SLPPRPP-----SP----- 225

QY 131 GROPQNVSELVGRPLPHPSYMRDVTYERHGGPLRYRRRPYPQEE---YLDIDOMFN- 185

DB 226 -RKPP-----PPPPAFMS-----SGSDYSDLPVLPSPSGLVLGESKSTFTY 270

QY 186 RELPOA-----SGLHNCSP-YKRGRLVTMTTARGLQSGDRATWFGLYYNIAGAF 238

DB 271 EELSRATNGSEANLIDQGGGTYHKGLPSGKEVAAVKOLKAG-----SQGGE 318

QY 239 FLHHVGLLEL--VNHKALDPAKWTIOKVFYQGRYVDSLADLEAFGLVNVVL--IPD 293

DB 319 REFOAEVEIISRVNHR-----HLVSLIGYCMAGVOYRLLVYEEFVN 358

QY 294 N-----GTGC---SWSLKSPPVPPRPPAPLQFYPOGFRFSVQSGRVAASLWTSFGLGA 343

DB 359 NNLEFHLHGKGRPTMEWSTRIKIALGSAKGLST-----LHE 394

QY 344 FSGPRIFDVRFQGERLYVEISLOEALAIYCGNSPAATTTRYVD---GGFG-----M 391

DB 395 DGNPKIITHROIKASNIILIDFEFAKVADFGAKIASDNTNHTVSTRWGCTGYLAPEYAA 454

QY 392 GKYYTP---LTRGVDCPYLAT-----YVDW----- 413

DB 455 GKLTESDVSFSGVVLLELTIGRRPVANNVYVDSLVDMARPLNRSAGEDEGLADS 514

QY 414 -----HFLLESQAPKTIIDAFV-----FEONGLP-----LRHHSDLY 448

DB 515 KMGNEIDREEMARVACAAACVRRSARRRPMGSIYVRLAEGNSVLSLIDNEGRPHGSHNVY 574

OY 449 SHYEGLAETLVNMSKSTLLNDYWDVTFHPSGAIETRFATGYISSAFLEGATGKYG 508
 DB 575 SSYGG-----STDY--DTSQYNDMDIKERKMAJGTOE-----YGTGEGYS 612
 OY 509 NOVSHTL 516
 DB 613 NPTSDYGL 620
 RESULT 14
 ABG13153
 ID ABG13153 standard; Protein; 446 AA.
 AC ABG13153;
 DT 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #13144.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder.
 OS Homo sapiens.
 PN WO200175067-A2.
 PD 11-OCT-2001.
 PF 30-MAR-2001; 2001WO-US08631.
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 PA (HYSE-) HYSQ INC.
 PI Drmanac RT, Liu C, Tang YT.
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS77340.
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 PS Claim 20; SEQ ID No 43512; 103pp; English.
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG0377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WRO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 446 AA;
 SQ
 Query Match 2.68; Score 106.5; DB 22; Length 446;
 Best Local Similarity 21.38; Pred. No. 0.24;

Matches 65; Conservative 39; Mismatches 98; Indels 103; Gaps 16;
 OY 352 VRFQGERLYEYI-SLQELALITGNSPAAATTRY-----VDGFGGKITTPLTRGV 402
 DB 211 LKAEGEGRAYQIMQLOKLVNMGCD---LTRYGOKVHKLTJHGFSGPNRDGTIGRG- 265
 OY 403 DCPY--LATYVDMHFLLESOAPKTIIDAFVCFQNGGLPLRRHSDL-----YSHFGG 454
 DB 266 GCTFCNVASFAD-----EAQHRISAEOLA-----HOANLVNRAKRYLGFQA 308
 OY 455 LAET---VLVRSMSYLLNDYWDVTFHPSGAIETRFATGYISSAF-----LEGAT 504
 DB 309 YSTFAEVOYLRSN-----YQAVEHHTGKVGITGCGYGGVMAAAYAPELACAV 360
 OY 505 GKRYNGVSEHTLGTHTSHAEKVDLDVAGLENNVMAEDMVFYPAVPPSPHQLORFY 564
 DB 361 PFYGRQAPTADVAKIEAPLLHFAELDTRINEGM-----PAYE-AALKR 403
 OY 565 TRKLEMEGAPLVSATPRYLXLSNHSNKKGHPRGYRIQMLSFAGEPLPONSSMARG 624
 DB 404 NNKVEE-----AYLPGVNH-----GRHNDSTPAYDKSAAD 434
 OY 625 FSWER 629
 DB 435 LAMOR 439
 RESULT 15
 AAE23407
 ID AAE23407 standard; Protein; 617 AA.
 AC AAE23407;
 DT 27-AUG-2002 (first entry)
 DE Lysine oxidase proenzyme mut3.
 KW Amino acid oxidase; insect infestation; gene therapy; boll weevil; BWV;
 KW corn rootworm; GRW; insecticide; wireworm; MW; Colorado potato beetle;
 KW CPB; Lysine oxidase; tedaralactam synthase; enzyme; E.C. 1.4.3.2;
 KM mutant; mutant.
 OS Unidentified.
 PS
 FH Location/Qualifiers
 FT Region .68..86
 FT /note- "Protease sensitive region"
 FT MISC-difference 70
 FT /note- "Wild-type Ala substituted with Gly"
 FT MISC-difference 71
 FT /note- "Wild-type Leu substituted with Gly"
 FT MISC-difference 72
 FT /note- "Wild-type Leu substituted with Gly"
 FT MISC-difference 73
 FT /note- "Wild-type Lys substituted with Gly"
 FT MISC-difference 74
 FT /label- "Unknown"
 FT MISC-difference 75
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 76
 FT /label- "Unknown"
 FT MISC-difference 77
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 78
 FT /label- "Unknown"
 FT MISC-difference 79
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 80
 FT /label- "Unknown"
 FT MISC-difference 81
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 82
 FT /label- "Unknown"
 FT MISC-difference 83
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 84
 FT /label- "Unknown"
 FT MISC-difference 85
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 86
 FT /label- "Unknown"
 FT MISC-difference 87
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 88
 FT /label- "Unknown"
 FT MISC-difference 89
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 90
 FT /label- "Unknown"
 FT MISC-difference 91
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 92
 FT /label- "Unknown"
 FT MISC-difference 93
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 94
 FT /label- "Unknown"
 FT MISC-difference 95
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 96
 FT /label- "Unknown"
 FT MISC-difference 97
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 98
 FT /label- "Unknown"
 FT MISC-difference 99
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 100
 FT /label- "Unknown"
 FT MISC-difference 101
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 102
 FT /label- "Unknown"
 FT MISC-difference 103
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 104
 FT /label- "Unknown"
 FT MISC-difference 105
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 106
 FT /label- "Unknown"
 FT MISC-difference 107
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 108
 FT /label- "Unknown"
 FT MISC-difference 109
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 110
 FT /label- "Unknown"
 FT MISC-difference 111
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 112
 FT /label- "Unknown"
 FT MISC-difference 113
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 114
 FT /label- "Unknown"
 FT MISC-difference 115
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 116
 FT /label- "Unknown"
 FT MISC-difference 117
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 118
 FT /label- "Unknown"
 FT MISC-difference 119
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 120
 FT /label- "Unknown"
 FT MISC-difference 121
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 122
 FT /label- "Unknown"
 FT MISC-difference 123
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 124
 FT /label- "Unknown"
 FT MISC-difference 125
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 126
 FT /label- "Unknown"
 FT MISC-difference 127
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 128
 FT /label- "Unknown"
 FT MISC-difference 129
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 130
 FT /label- "Unknown"
 FT MISC-difference 131
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 132
 FT /label- "Unknown"
 FT MISC-difference 133
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 134
 FT /label- "Unknown"
 FT MISC-difference 135
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 136
 FT /label- "Unknown"
 FT MISC-difference 137
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 138
 FT /label- "Unknown"
 FT MISC-difference 139
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 140
 FT /label- "Unknown"
 FT MISC-difference 141
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 142
 FT /label- "Unknown"
 FT MISC-difference 143
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 144
 FT /label- "Unknown"
 FT MISC-difference 145
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 146
 FT /label- "Unknown"
 FT MISC-difference 147
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 148
 FT /label- "Unknown"
 FT MISC-difference 149
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 150
 FT /label- "Unknown"
 FT MISC-difference 151
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 152
 FT /label- "Unknown"
 FT MISC-difference 153
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 154
 FT /label- "Unknown"
 FT MISC-difference 155
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 156
 FT /label- "Unknown"
 FT MISC-difference 157
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 158
 FT /label- "Unknown"
 FT MISC-difference 159
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 160
 FT /label- "Unknown"
 FT MISC-difference 161
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 162
 FT /label- "Unknown"
 FT MISC-difference 163
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 164
 FT /label- "Unknown"
 FT MISC-difference 165
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 166
 FT /label- "Unknown"
 FT MISC-difference 167
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 168
 FT /label- "Unknown"
 FT MISC-difference 169
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 170
 FT /label- "Unknown"
 FT MISC-difference 171
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 172
 FT /label- "Unknown"
 FT MISC-difference 173
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 174
 FT /label- "Unknown"
 FT MISC-difference 175
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 176
 FT /label- "Unknown"
 FT MISC-difference 177
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 178
 FT /label- "Unknown"
 FT MISC-difference 179
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 180
 FT /label- "Unknown"
 FT MISC-difference 181
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 182
 FT /label- "Unknown"
 FT MISC-difference 183
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 184
 FT /label- "Unknown"
 FT MISC-difference 185
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 186
 FT /label- "Unknown"
 FT MISC-difference 187
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 188
 FT /label- "Unknown"
 FT MISC-difference 189
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 190
 FT /label- "Unknown"
 FT MISC-difference 191
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 192
 FT /label- "Unknown"
 FT MISC-difference 193
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 194
 FT /label- "Unknown"
 FT MISC-difference 195
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 196
 FT /label- "Unknown"
 FT MISC-difference 197
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 198
 FT /label- "Unknown"
 FT MISC-difference 199
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 200
 FT /label- "Unknown"
 FT MISC-difference 201
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 202
 FT /label- "Unknown"
 FT MISC-difference 203
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 204
 FT /label- "Unknown"
 FT MISC-difference 205
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 206
 FT /label- "Unknown"
 FT MISC-difference 207
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 208
 FT /label- "Unknown"
 FT MISC-difference 209
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 210
 FT /label- "Unknown"
 FT MISC-difference 211
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 212
 FT /label- "Unknown"
 FT MISC-difference 213
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 214
 FT /label- "Unknown"
 FT MISC-difference 215
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 216
 FT /label- "Unknown"
 FT MISC-difference 217
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 218
 FT /label- "Unknown"
 FT MISC-difference 219
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 220
 FT /label- "Unknown"
 FT MISC-difference 221
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 222
 FT /label- "Unknown"
 FT MISC-difference 223
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 224
 FT /label- "Unknown"
 FT MISC-difference 225
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 226
 FT /label- "Unknown"
 FT MISC-difference 227
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 228
 FT /label- "Unknown"
 FT MISC-difference 229
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 230
 FT /label- "Unknown"
 FT MISC-difference 231
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 232
 FT /label- "Unknown"
 FT MISC-difference 233
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 234
 FT /label- "Unknown"
 FT MISC-difference 235
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 236
 FT /label- "Unknown"
 FT MISC-difference 237
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 238
 FT /label- "Unknown"
 FT MISC-difference 239
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 240
 FT /label- "Unknown"
 FT MISC-difference 241
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 242
 FT /label- "Unknown"
 FT MISC-difference 243
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 244
 FT /label- "Unknown"
 FT MISC-difference 245
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 246
 FT /label- "Unknown"
 FT MISC-difference 247
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 248
 FT /label- "Unknown"
 FT MISC-difference 249
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 250
 FT /label- "Unknown"
 FT MISC-difference 251
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 252
 FT /label- "Unknown"
 FT MISC-difference 253
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 254
 FT /label- "Unknown"
 FT MISC-difference 255
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 256
 FT /label- "Unknown"
 FT MISC-difference 257
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 258
 FT /label- "Unknown"
 FT MISC-difference 259
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 260
 FT /label- "Unknown"
 FT MISC-difference 261
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 262
 FT /label- "Unknown"
 FT MISC-difference 263
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 264
 FT /label- "Unknown"
 FT MISC-difference 265
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 266
 FT /label- "Unknown"
 FT MISC-difference 267
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 268
 FT /label- "Unknown"
 FT MISC-difference 269
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 270
 FT /label- "Unknown"
 FT MISC-difference 271
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 272
 FT /label- "Unknown"
 FT MISC-difference 273
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 274
 FT /label- "Unknown"
 FT MISC-difference 275
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 276
 FT /label- "Unknown"
 FT MISC-difference 277
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 278
 FT /label- "Unknown"
 FT MISC-difference 279
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 280
 FT /label- "Unknown"
 FT MISC-difference 281
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 282
 FT /label- "Unknown"
 FT MISC-difference 283
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 284
 FT /label- "Unknown"
 FT MISC-difference 285
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 286
 FT /label- "Unknown"
 FT MISC-difference 287
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 288
 FT /label- "Unknown"
 FT MISC-difference 289
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 290
 FT /label- "Unknown"
 FT MISC-difference 291
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 292
 FT /label- "Unknown"
 FT MISC-difference 293
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 294
 FT /label- "Unknown"
 FT MISC-difference 295
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 296
 FT /label- "Unknown"
 FT MISC-difference 297
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 298
 FT /label- "Unknown"
 FT MISC-difference 299
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 300
 FT /label- "Unknown"
 FT MISC-difference 301
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 302
 FT /label- "Unknown"
 FT MISC-difference 303
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 304
 FT /label- "Unknown"
 FT MISC-difference 305
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 306
 FT /label- "Unknown"
 FT MISC-difference 307
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 308
 FT /label- "Unknown"
 FT MISC-difference 309
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 310
 FT /label- "Unknown"
 FT MISC-difference 311
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 312
 FT /label- "Unknown"
 FT MISC-difference 313
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 314
 FT /label- "Unknown"
 FT MISC-difference 315
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 316
 FT /label- "Unknown"
 FT MISC-difference 317
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 318
 FT /label- "Unknown"
 FT MISC-difference 319
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 320
 FT /label- "Unknown"
 FT MISC-difference 321
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 322
 FT /label- "Unknown"
 FT MISC-difference 323
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 324
 FT /label- "Unknown"
 FT MISC-difference 325
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 326
 FT /label- "Unknown"
 FT MISC-difference 327
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 328
 FT /label- "Unknown"
 FT MISC-difference 329
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 330
 FT /label- "Unknown"
 FT MISC-difference 331
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 332
 FT /label- "Unknown"
 FT MISC-difference 333
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 334
 FT /label- "Unknown"
 FT MISC-difference 335
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 336
 FT /label- "Unknown"
 FT MISC-difference 337
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 338
 FT /label- "Unknown"
 FT MISC-difference 339
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 340
 FT /label- "Unknown"
 FT MISC-difference 341
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 342
 FT /label- "Unknown"
 FT MISC-difference 343
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 344
 FT /label- "Unknown"
 FT MISC-difference 345
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 346
 FT /label- "Unknown"
 FT MISC-difference 347
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 348
 FT /label- "Unknown"
 FT MISC-difference 349
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 350
 FT /label- "Unknown"
 FT MISC-difference 351
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 352
 FT /label- "Unknown"
 FT MISC-difference 353
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 354
 FT /label- "Unknown"
 FT MISC-difference 355
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 356
 FT /label- "Unknown"
 FT MISC-difference 357
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 358
 FT /label- "Unknown"
 FT MISC-difference 359
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 360
 FT /label- "Unknown"
 FT MISC-difference 361
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 362
 FT /label- "Unknown"
 FT MISC-difference 363
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 364
 FT /label- "Unknown"
 FT MISC-difference 365
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 366
 FT /label- "Unknown"
 FT MISC-difference 367
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 368
 FT /label- "Unknown"
 FT MISC-difference 369
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 370
 FT /label- "Unknown"
 FT MISC-difference 371
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 372
 FT /label- "Unknown"
 FT MISC-difference 373
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 374
 FT /label- "Unknown"
 FT MISC-difference 375
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 376
 FT /label- "Unknown"
 FT MISC-difference 377
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 378
 FT /label- "Unknown"
 FT MISC-difference 379
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 380
 FT /label- "Unknown"
 FT MISC-difference 381
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 382
 FT /label- "Unknown"
 FT MISC-difference 383
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 384
 FT /label- "Unknown"
 FT MISC-difference 385
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 386
 FT /label- "Unknown"
 FT MISC-difference 387
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 388
 FT /label- "Unknown"
 FT MISC-difference 389
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 390
 FT /label- "Unknown"
 FT MISC-difference 391
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 392
 FT /label- "Unknown"
 FT MISC-difference 393
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 394
 FT /label- "Unknown"
 FT MISC-difference 395
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 396
 FT /label- "Unknown"
 FT MISC-difference 397
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 398
 FT /label- "Unknown"
 FT MISC-difference 399
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 400
 FT /label- "Unknown"
 FT MISC-difference 401
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 402
 FT /label- "Unknown"
 FT MISC-difference 403
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 404
 FT /label- "Unknown"
 FT MISC-difference 405
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 406
 FT /label- "Unknown"
 FT MISC-difference 407
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 408
 FT /label- "Unknown"
 FT MISC-difference 409
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 410
 FT /label- "Unknown"
 FT MISC-difference 411
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 412
 FT /label- "Unknown"
 FT MISC-difference 413
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 414
 FT /label- "Unknown"
 FT MISC-difference 415
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 416
 FT /label- "Unknown"
 FT MISC-difference 417
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 418
 FT /label- "Unknown"
 FT MISC-difference 419
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 420
 FT /label- "Unknown"
 FT MISC-difference 421
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 422
 FT /label- "Unknown"
 FT MISC-difference 423
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 424
 FT /label- "Unknown"
 FT MISC-difference 425
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 426
 FT /label- "Unknown"
 FT MISC-difference 427
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 428
 FT /label- "Unknown"
 FT MISC-difference 429
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 430
 FT /label- "Unknown"
 FT MISC-difference 431
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 432
 FT /label- "Unknown"
 FT MISC-difference 433
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 434
 FT /label- "Unknown"
 FT MISC-difference 435
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 436
 FT /label- "Unknown"
 FT MISC-difference 437
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 438
 FT /label- "Unknown"
 FT MISC-difference 439
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 440
 FT /label- "Unknown"
 FT MISC-difference 441
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 442
 FT /label- "Unknown"
 FT MISC-difference 443
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 444
 FT /label- "Unknown"
 FT MISC-difference 445
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 446
 FT /label- "Unknown"
 FT MISC-difference 447
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 448
 FT /label- "Unknown"
 FT MISC-difference 449
 FT /note- "Wild-type Arg substituted with Xaa"
 FT MISC-difference 450
 FT /label- "Unknown"
 FT MISC-difference 451
 FT /note- "Wild-type Glu substituted with Xaa"
 FT MISC-difference 452
 FT /label- "Unknown"
 FT MISC-difference 453
 FT /note- "Wild-type Ala substituted with Xaa"
 FT MISC-difference 454
 FT /label- "Unknown"
 FT MISC-difference 455
 FT /note- "Wild-type Arg substituted with Xaa"<


```

Query Match          2.5%; Score 102.5; DB 4; Length 541.
Best Local Similarity 19.6%; Pred. No. 0.14;
Matches 109; Conservative 60; Mismatches 205; Indels 183; Gaps 22

QY 170 PVLQGEYLDIDQMIFNRRELPOASGLLHNCFF---YKRGRLVMTTAPRGISGDRAT 225
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 40 PELVQHAEDMHEWING-NLRAGATYQTCIFAVPVGARNG-GLVYTCQPRNLEHLKKAR 97
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 226 -----WFLYLYNISAGFF-----LHHVGLLELLVAKALDPAKWTIOKVY 266
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 98 PDNYKGPFWHCVFRLDLDDGIFNSDGDWTLAQRKTAALFTTTLRTLTASMSRVSRSKI-- 155
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 267 QGRYYSLAQL---EQAGEGLVNVYLLFPDNGTGSWSLSKSPVPPARPPLQFYPGGPRF 323
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 156 HORLLPILDLDAKKAQAYVDLDDLRLTLFFDNLGCLAFG-----KQETLLAGGLPNEFA 209
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 324 SVQGSRVASSLWTFSF-----GLGAFSGPRIFDVROGERLYEISIQEALATY 372
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 210 SAFPDATEYTLRLRFPEFLMRCKKMLGLG-----METLTSSMAVY 251
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 373 GGNSPAAMTTRYVDGFGMGKYYTPLYR-----GYDC 404
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 252 DQYLAAYIKRKLLELAAGNGKCDTATNHDLLSRFMKSGSYDESLOHVALNPILAGRGT 311
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 405 PLYLATYVDMHFLLESCAP---KTIIDAFQVEQONGL-----PLRRHSDLYSHY 451

```

Query Match:	2.5%	Score 102.5	DB 4	Length 541	
Best Local Similarity	19.6%	Pred. No. 0.14			
Matches 109	Conservative	60	Mismatches 205	Indels 183	
			Gaps	22	
Qy	170	PVLFOEYLDIDOMIENRELPOASGLLHHCFF----	YKHGRMLVMTTAPRGLOSGDRAT	225	
Db	40	PLGVGHARDMEHWING-NLRAGSGYGCIFAVPGVARG	-GLVYITOPRNLHENLKAY	97	
Qy	226	-----WGLYYNIGAGFF-----	LHHVGLLELLVNHKALDPAMWTIOKAY	266	
Db	98	FDNYKGFPMHGVFEDLLDGDGFNSDGDITWLAKRTA	LEFTTRLTASRWVSKSI--	155	
Qy	267	QGRYVDSLALQ---	RAOPFAGLVNVLLPDNCTGSGMSLSKSPVPGAPAPLQFYPOGPRF	323	
Db	156	HGRLLPILDAKKAQYVDLDLRLTFDNLGIAFG----	KDFELAOLGLENENA	209	
Qy	324	SVQSGRVASLWTFSG-----	GLGAFSGPRIDFVROGERLYEISLOEALAY	372	
Db	210	SAFDSATETLNRFLFPERLMRCKKMLIG-----	WETTLTSMANV	251	
Qy	373	GNNSPAAMTTRYVDGFGNGKTYTPLTR-----	-----GVDC	404	
Db	252	DOYLAIVIKRKLTELAANGKCDTAAITHDDLSPRMKSG	YSDESIQHVALNFIILAGRPT	311	
Qy	405	PYLATYVNMHFLFESQAR----	KTRDAFCVPEQOGL-----	PLRRHSDLSHY	451
Db	312	SSVA--LSWFELVSTHRAVERKIYRELCSYLAASRGA	HPALMLAEPTFEELDLVYL	369	


```

OY 452 FGGAEATLVVRSM-----STLTANTDYAMDVFEHSG-AIEIRFATQYISSAFLEFGATG 506
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 370 KAALSETIRLPSYPEDSKHVVADDTLPDGTFFPAGSSVITYISAGRMKG----- 420
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 507 YGNVSEHTTGTVHTSHAFKVDLDVAGLEMMVAEOMVEFPAVAPSPPEHOLQLOVTR 566
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 421 -----VMGED-----CLERPE---RWLSADG 439
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 567 KLEMEQAAFLVGSATPRVITYLLASNHSNNKGHPGRGRIOMLSPAGELPDONSMAFGFS 626
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 440 TKFEQHSYKRVAAVNAAGRVCL-----GKDLAY-LQMKINAGSVLLRH----- 481
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 627 MERYOLAVTORKEEPS 643
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 482 --RLTVAPGRHVEQMS 496
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT 4

US-09-063-733A-46
Sequence 46, Application US/09063733A
Patent No. 6372211
GENERAL INFORMATION:
APPLICANT: Isaac, Barbara G.
APPLICANT: Greentplate, John T.
APPLICANT: Purcell, John P.
APPLICANT: Romano, Charles P.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROLLING
TITLE OF INVENTION: INSECTS
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold White & Durkee
STREET: PO Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/063,733A
FILING DATE: 21-APR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Patterson, Melinda L.
REGISTRATION NUMBER: 33,062
REFERENCE/DOCKET NUMBER: MOBT-022
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-787-1400
TELEFAX: 713-787-1440
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 617 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
US-09-063-733A-46

```

Query Match      2.5%; Score 101; DB 4; Length 617;
Best Local Similarity 23.3%; Pred. No. 0.25;
Matches 81; Conservative 39; Mismatches 138; Indels 90; Gaps 19;

QY 63 LSREELTAVMRFLORLG--PGLYDAQAR-----SDNCVFSEVLQPPKAA 109
      | | | | | : | | | | | | | | | | | | | | | | | | | | |
Db 18 LIREKAKELNITTELGEVPGIIPNERERFLGGYSHDNLPSDLYSSI-----KPL 71

QY 110 LAHLDRGSP--PAREALAIYFEGROPENASELVYG-----PLPHEPMDRVTERHG 161
      | | | | | : | | | | | | | | | | | | | | | | | | | | |
Db 72 LKEAPAEELPPRK--VCIIVAG-----VSLYIAMIIDIKITPNLYIDIFESSRKG 123

```

```

OY 162 GRLPYHRRPVLFOCYLIDIDOMIFRELPOQASGLHHCCFCFKHNSRNLVYMTAPRGLOSG 221
Db 124 GRLYTHHTDAKKHDYIDIGAMRY----PDLPMSKRTNFKRRTSGMPLI-----KYLIDG 173
OY 222 DRATPFGLYYNISGAFGEFLHHVGLLELVN-----HKALDPARMTIQKVFYOGGRYDS 273
Db 174 ENTP--QLYNN-----HFFAKGVSDPYMWVSANGTVPDDVDVDSYGEKLQOAF--GYKKEK 225
OY 274 LAOLEAOFEACLVANVLIPDNGTGSNSLKSVPVPGCAPPLQIFPOQPRFSVQGSRASS 333
Db 226 LAE---DFDKCFDEIMLVDMDTT--REYLKKGSGKGEAPKXDF-----FAIQMMENTON- 273
OY 334 LMTFSFGAFCGSGRIPFVRQOGERLVEISLO-----EALAIYCGNS 376
Db 274 -----TGNTLFPDQAF--SESVIDSPFDNPMTKPKMYCIGGGS 309

```

RESULT 5

US-08-506-2968-14
Sequence 14, Application US/08506296B
Patent No. 6313265
GENERAL INFORMATION:
APPLICANT: Phillips, Greg
APPLICANT: Cunningham, Bruce A.
APPLICANT: Crossin, Kathryn L.
TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES
TITLE OF INVENTION: CONTAINING FIBROBLAST TYPE III REPEATS AND METHODS OF USE
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESS: The Scripps Research Institute
STREET: 10550 No. 6313265th Torrey Pines Road, TPC-8
CITY: La Jolla
STATE: California
COUNTRY: U.S.
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/506,296B
FILING DATE: 24-JUL-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: TSRI 488.0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 554-2937
TELEFAX: (619) 554-6312
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1253 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-506-2968-14

```

Query Match 2.5%; Score 100.5; DB 4; Length 1253;
Best Local Similarity 18.5%; Pred. No. 0.95;
Matches 160; Conservative 89; Mismatches 269; Indels 349; Gaps 40

QY 23 VLVNKGCG-----GERSQLPHCPSPSPSAQ-WTPHGSQLFA----- 61
      :||| |||
Db 299 LKVGEDDEGERCLAENSLGSAHAYYYTVEAPYMLKPOSHLGPETARLDCQVQG 358
      :|||
QY 62 -----DLSRELLAVMRFILQRLGPGVLDAQAQRPSPNCFSVE----- 100
      :|||
Db 359 RPQPEVTWRINGIPVEELAKDQKYRQ---GALLISNVPQSPDTWTOGEARNRHGLLLA 415
      :|||
QY 101 -----LQPPKAAALAHLDKGSPPAREALALVFEGGRQPPQPVSLVVGRLPHRPYMD 154
      :||| :|||

```

```

Db 416 NAVIYVQLPAKILITADNOTYMAVOGSTAYLLCKAFG-APVPSY----- 458
OY 155 VYVERHGGPLPYHRRPVLFQBYLDIDOMIF---NRELPOASGLLHCCFYKHKRNLVTM 211
Db 459 -----QWLEDEGTVLQDERFFPYANG----- 480
OY 212 TTAARGLOSDBRATWGLLYNIGAGFFLHVGLLELVNKAALDPAWTKQVYQGRY 271
Db 481 TLGIRLDQANDTCRYFCLANDON-----NNTIMANKVKDATIT----- 521
OY 272 DSLAQLAEQAEGLVNVVLIIPDNGTGSWSLSKSPVPPAPPLQFYQGFPSVO---GSR 329
Db 522 -----QGRSTLEKKGR 534
OY 330 VASSLMTFFSGGLAFSGPRIFD-----VRQGE-RLVYEISLOEALATYGCNSPAAWTT 382
Db 535 V-----TFTCOASFDPISLPISITWRGDRDLOELGSDKFIEDGRL-VIHSL 581
OY 383 RYVDGFGMGKYYTPLRGRDCPYLATYVDMHFLLESQA-----PKTRDAFCV 431
Db 582 DYSD-----QGNYSQVASTEID-----VESRAQLLVYSGPVPRLVLSDLHL 625
OY 432 FEONO-----GLPLRRHSDLYSHYFGGLAETVLVVRSMSTLLNDYVDTVFHPSGAIEI 487
Db 626 LTQOVAVSMSPADHNAPIEK-----YDIEFEDKEMAPE 660
OY 488 RFIYATG-----YISSAFLFGATGKYG---NQVSEHLGTYHT-HSAHF 527
Db 661 KWSYLGKVPNGQSTTKLSPIYHYTFRVTAINKYGGPEPVSBE---TVTPEAPEK 716
OY 528 VDLVAGLEMMVADNVFVMAVWSP-----EHQLQ-RLQVTRKLEMEQA 575
Db 717 NPVYVKG-----EGNETTMVITWKPLRMMDNNAPOVYRVWRQGTG--PMQOI 767
OY 576 A---FLVGSATPRLYLASNHNKMGHPRGYRIQMLSPAGE-PLPQNSMARGESWERYQ 631
Db 768 VSDPFLVSNSTFV-----PYEIKVOAVNSQKGPQO---VTIGXGSEDP 812
OY 632 LAVQKREEESSSVFNQNDPMAPTVDFSPINNETIAGDLYAMVYTAG-----FLH 684
Db 813 QATPELEIGETILNSSAVLYK--WRP-VDLAOKGH--LRGINVYWRGSGRKHSHRIH 867
OY 685 IPHAEDIPNTVTNGVGFPLRPYNFDEDEPSFSADSIYRGDODA-----G 732
Db 868 KDHVVPANTSV---ILSGLRPSSYLEVOAENG-----RGSPPAEFTFSPREGVPG 919
OY 733 ACEVNPPLACLPQAAACAPDLPAFESHG 759
Db 920 HPEALHLECOGNTSILLRMQPLSHNG 946

RESULT 6
US-09-314-242-2
: Sequence 2, Application US/09314242A
: Patent No. 6248575
: GENERAL INFORMATION:
: APPLICANT: Elizabeth J. Gollightly
: TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
: FILE OF INVENTION: Having L-Amino Acid Oxidase Activity
: CURRENT APPLICATION NUMBER: US/09/314,242A
: EARLIER FILING DATE: 1999-05-18
: EARLIER APPLICATION NUMBER: 09/080,428
: NUMBER OF SEQ ID NOS: 12
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 617
: TYPE: PRT
: ORGANISM: Trichoderma harzianum
US-09-314-242-2
Query Match 2.4%; Score 100; DB 4; Length 617;

```

```

Best Local Similarity 23.9%; Pred. No. 0.31;
Matches 83; Conservative 36; Mismatches 139; Indels 90; Gaps 18;

OY 63 LSREELTAVNRFLTQRLG--PGLVDAQAAR-----PSDNCVSYELQLPKAA 109
Db 18 LIREKVAKEINILTERGEVPIPPNEGRLGGYSHDMLPSDPLYSI-----KPAL 71
OY 110 LAHLDGRSP--PPAREALATVFGROPQPVNSELVYG-----PLPSPYMRDVTYVRHG 161
Db 72 LKEAPRAEELPPRK--VCIVGAG-----VSGLYIMILDDLKIPULTYDISSSRTG 123
OY 162 GLPYHRRPVLFQBYLDIDOMIFNRELPOASGLLHCCFYKHKRNLVTTAARGLOSG 221
Db 124 GRLYTHFTDAKHDIYIGAMRY-PDIPSMK-----RFNLFKTKKPLIYYL 171
OY 222 DRATWFLGLYINISAGFFLHVGLLELVN-----HKALDPAWTKQVYQGRYSD 273
Db 172 DGEWTPOLYNN---HFAKGVSDPYVAVANGSTVPDDVDVDSVGERLQOAF--GYKEK 225
OY 274 LAQLEAQFAGLVNVVLIIPDNGTGSWSLSKSPVPPAPPLQFYQGFPSVQSSRASS 333
Db 226 LAE---DFDKGFDELMLVDWTT--REYLKRGKGPGEAPKYDF-----FALQMETON- 273
OY 334 LMTFSGLGAFSGPRIFDVRFQGERLYEISLQ-----EALAIYGNS 376
Db 274 -----TGYNLFQDAF-SESVIDSPFDNPTKPMYCIEGTS 309

RESULT 7
US-09-663-733A-57
: Sequence 57, Application US/09063733A
: Patent No. 6372211
: GENERAL INFORMATION:
: APPLICANT: Isaac, Barbara G.
: APPLICANT: Greenplate, John T.
: APPLICANT: Purcell, John P.
: APPLICANT: Romano, Charles P.
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROLLING
: NUMBER OF SEQUENCES: 58
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Arnold White & Durkee
: STREET: PO Box 4433
: CITY: Houston
: STATE: TX
: COUNTRY: USA
: ZIP: 77210-4433
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/063,733A
: FILING DATE: 21-APR-1998
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Paterson, Melinda L.
: REGISTRATION NUMBER: 33,062
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 713-787-1400
: TELEFAX: 713-787-1440
: INFORMATION FOR SEQ ID NO: 57:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 617 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: linear
US-09-063-733A-57
Query Match 2.3%; Score 95; DB 4; Length 617;
Best Local Similarity 22.5%; Pred. No. 1;

```


COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/445,464C
FILING DATE: 22-MAY-1995

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/044,857
FILING DATE: 08-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Ruzsala, Lois K.
REGISTRATION NUMBER: 39,074
REFERENCE/DOCKET NUMBER: BEH-7261 DIV 1
TELECOMMUNICATION INFORMATION:

TELEPHONE: (847) 267-5364

TELEFAX: (847) 267-5376

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 486 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-08-445-464C-8

Query Match

Best Local Similarity 20.8%; Score 93.5; DB 3; Length 486;

Matches 81; Conservative 61; Mismatches 151; Indels 97; Gaps 20;

```

QY 406 YLATYVDHMF-----LLESQAPKTIRDAFCVEONQGLPLRRHSDLSHYFGCLA 456
DB 31 YKKGILQEHFAIYGTARQOLSDDEFKQVYDSIKDFTEDA-----QAEAFIAHFSYRA 84
QY 457 ETVLVVRSNSTL-----LNTDYWDVTFHPSGAIEIRFYAT--GYISSAFLFGATGK 506
DB 85 HDVTDAAASYGILKSAIEEATKFDIDGNRIFYMS--VAPRFEGTIAKYLKSEGLAETG- 141
QY 507 YGNQVSEHTLGYVTHSAHFKVDLDVAGLENTWV-----AEDMVFPMAV----- 551
DB 142 YNRLIMEKPEFGTSYATAEELQSDLENAFDDQFLRIDHYLGKEVONIAALRFGNPIFDA 201
QY 552 PMSPEHOLRLQVY--RKLLEMEBOAFLVGSATPRYLYLASNSHKNKGHPRGYRIOMLSF 610
DB 202 AMNKDY--IKNVQVTLAEVLGVEERAGY--DTGALLDMQNH-----MQLVGM 248
QY 611 AGEPLPQ--NSSMARGFSWERYQLAVTORKEEPESSSVFNQNDPMAPYVESDFT---- 664
DB 249 LAMEKPESEFNDKDIRAKKNAFN--ALKIYNEEVNKYFVRAOYGA--GDTADYKPYLEAD 306
QY 665 -----NNETIAGK---DLYAVYTAGFLHPIHADIENRYTVGN-----GYGFELR 706
DB 307 VPADSKNNTFIAGELQDLPLRW-----EGVFPYVSGRLAAKOTRVDIVFKAG 355
QY 707 PYNFDEDPSEFYSADSIYFRGDODAGACEV 736
DB 356 TPNFGSEQEAQESVLSITIT--DPKGAIEL 382

```

RESULT 15

US-08-044-857D-8

Sequence 8, Application US/08044857D

Patent No. 6453288

GENERAL INFORMATION:

APPLICANT: Jakobovits, Edward B.

Silen, Joy L.

Levy, Mark J.

Goodman, Thomas C.

Becker, Martin

Caldwell, Robert M.

Bott, Richard R.

Barnett, Christopher C.

Glucose-6-Phosphate Dehydrogenases

NUMBER OF SEQUENCES: 124

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dade Behring Inc.

STREET: 1717 Deerfield Road

CITY: Deerfield

STATE: Illinois

COUNTRY: USA

ZIP: 60015

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/044,857D

FILING DATE: 08-APR-1993

CLASSIFICATION: <unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Ruzsala, Lois K.

REGISTRATION NUMBER: 39,074

REFERENCE/DOCKET NUMBER: BEH-7261

TELECOMMUNICATION INFORMATION:

TELEPHONE: (847) 267-5364

TELEFAX: (847) 267-5376

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 486 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Leuconostoc dextranctum

STRAIN: ATCC 19255

SEQUENCE DESCRIPTION: SEQ ID NO: 8:

US-08-044-857D-8

Query Match

Best Local Similarity 20.8%; Score 93.5; DB 4; Length 486;

Matches 81; Conservative 61; Mismatches 151; Indels 97; Gaps 20;

```

QY 406 YLATYVDHMF-----LLESQAPKTIRDAFCVEONQGLPLRRHSDLSHYFGCLA 456
DB 31 YKKGILQEHFAIYGTARQOLSDDEFKQVYDSIKDFTEDA-----QAEAFIAHFSYRA 84
QY 457 ETVLVVRSNSTL-----LNTDYWDVTFHPSGAIEIRFYAT--GYISSAFLFGATGK 506
DB 85 HDVTDAAASYGILKSAIEEATKFDIDGNRIFYMS--VAPRFEGTIAKYLKSEGLAETG- 141
QY 507 YGNQVSEHTLGYVTHSAHFKVDLDVAGLENTWV-----AEDMVFPMAV----- 551
DB 142 YNRLIMEKPEFGTSYATAEELQSDLENAFDDQFLRIDHYLGKEVONIAALRFGNPIFDA 201
QY 552 PMSPEHOLRLQVY--RKLLEMEBOAFLVGSATPRYLYLASNSHKNKGHPRGYRIOMLSF 610
DB 202 AMNKDY--IKNVQVTLAEVLGVEERAGY--DTGALLDMQNH-----MQLVGM 248
QY 611 AGEPLPQ--NSSMARGFSWERYQLAVTORKEEPESSSVFNQNDPMAPYVESDFT---- 664
DB 249 LAMEKPESEFNDKDIRAKKNAFN--ALKIYNEEVNKYFVRAOYGA--GDTADYKPYLEAD 306
QY 665 -----NNETIAGK---DLYAVYTAGFLHPIHADIENRYTVGN-----GYGFELR 706
DB 307 VPADSKNNTFIAGELQDLPLRW-----EGVFPYVSGRLAAKOTRVDIVFKAG 355
QY 707 PYNFDEDPSEFYSADSIYFRGDODAGACEV 736

```

Db 356 TFFGSEGEAQESVLSII---DPKGAIEL 382

Search completed: May 20, 2003, 11:40:55
Job time : 36.5906 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:37:59 ; Search time 121.567 Seconds

(without alignments)
622.443 Million cell updates/sec

Title: US-10-081-408-2

Perfect score: 4095
Sequence: 1 MNOKTIVLLILAVITIFAL.....QAAACAPDLPAFSGHGSFN 763

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 9917265 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PTI_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/PTI_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/METHOD_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4095	100.0	763	9	US-10-081-408-2
2	3967	96.9	729	9	US-10-081-408-20
3	2536.5	61.9	928	10	US-09-919-497-51
4	106	2.6	617	9	US-10-005-530-56
5	104	2.5	980	9	US-09-908-193-22
6	103	2.5	2224	9	US-10-115-563-14
7	102	2.5	1194	9	US-09-738-626-5832
8	101	2.5	617	9	US-10-005-530-46
9	99	2.4	944	9	US-10-213-990-27
10	95.5	2.3	581	9	US-09-738-626-4914
11	95	2.3	617	9	US-10-002-530-57
12	93.5	2.3	878	10	US-09-912-020-364
13	93	2.3	480	10	US-09-734-569-172
14	93	2.3	601	10	US-09-925-501-844
15	93	2.3	1447	1	US-08-954-701A-19
16	93	2.3	1447	10	US-09-898-533-5
17	92.5	2.3	635	9	US-10-099-895-33
18	92	2.2	904	9	US-09-712-363-247
19	91.5	2.2	994	9	US-10-004-551-16

20	91	2.2	390	9	US-09-738-626-6558	Sequence 6558, Ap
21	91	2.2	1247	9	US-09-908-193-2	Sequence 2, Appl
22	91	2.2	2507	9	US-09-819-104A-2	Sequence 2, Appl
23	90.5	2.2	388	9	US-09-361-630-1	Sequence 1, Appl
24	90	2.2	589	10	US-09-740-041-2	Sequence 2, Appl
25	90	2.2	617	9	US-10-005-530-56	Sequence 56, Appl
26	89.5	2.2	306	9	US-10-017-910-7	Sequence 71, Appl
27	89.5	2.2	630	9	US-10-101-464A-71	Sequence 4, Appl
28	89.5	2.2	911	9	US-09-855-754-4	Sequence 2, Appl
29	89.5	2.2	911	9	US-10-227-353-4	Sequence 4, Appl
30	89.5	2.2	953	10	US-09-845-583-4	Sequence 2, Appl
31	89.5	2.2	3695	9	US-10-037-182-2	Sequence 2, Appl
32	89	2.2	365	9	US-09-925-299-1007	Sequence 1007, Ap
33	89	2.2	365	10	US-09-925-299-1007	Sequence 1007, Ap
34	89	2.2	1430	9	US-10-087-993-34	Sequence 34, Appl
35	88.5	2.2	583	9	US-09-942-185-2	Sequence 2, Appl
36	88.5	2.2	616	10	US-09-925-300-1519	Sequence 1519, Ap
37	88.5	2.2	882	9	US-10-174-580-574	Sequence 574, App
38	88.5	2.2	882	9	US-10-175-758-574	Sequence 574, App
39	88.5	2.2	882	9	US-10-175-737-574	Sequence 574, App
40	88.5	2.2	882	9	US-10-173-706-574	Sequence 574, App
41	88.5	2.2	882	9	US-10-175-738-574	Sequence 574, App
42	88.5	2.2	882	9	US-10-175-738-574	Sequence 574, App
43	88.5	2.2	882	9	US-10-176-482-574	Sequence 574, App
44	88.5	2.2	882	9	US-10-176-757-574	Sequence 574, App
45	88.5	2.2	882	9	US-10-176-913-574	Sequence 574, App

ALIGNMENTS

RESULT 1
US-10-081-408-2 Application US/10081408
; Sequence 2, US/10081408
; Patent No. US20020160482A1
; GENERAL INFORMATION:
; APPLICANT: Abtams, n, Lars
; TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
; FILE REFERENCE: 13425-053001
; CURRENT FILING DATE: 2001-02-23
; PRIOR FILING DATE: 2001-02-23
; CURRENT APPLICATION NUMBER: US/10/081,408
; PRIOR APPLICATION NUMBER: SE 0100625-3
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 763
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-081-408-2

Query Match 100.0%; Score 4095; DB 9; Length 763;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 763; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MNOKTIVLLILAVITIFALVGVLRGDDGGEPSQLPHCSVSPSAQPTNHPGOSOLF	60
DB	1	MNOKTIVLLILAVITIFALVGVLRGDDGGEPSQLPHCSVSPSAQPTNHPGOSOLF	60
QY	61	ADLSREELTAVVRFLLTORGLVDAQAQRPDNCVFSELDLPKAAALAILDGSPPP	120
DB	61	ADLSREELTAVVRFLLTORGLVDAQAQRPDNCVFSELDLPKAAALAILDGSPPP	120
QY	121	AREALAIYFFGQPOPNSVGVPLPHPSYRDVTVERRHGPPLVHRRPVLFQEYLDID	180
DB	121	AREALAIYFFGQPOPNSVGVPLPHPSYRDVTVERRHGPPLVHRRPVLFQEYLDID	180
QY	181	QMIFRRELPOAGSLHHCFCYKRRGRNLVTMTARGLDGSGRATWFGLYNINSAGFFL	240
DB	181	QMIFRRELPOAGSLHHCFCYKRRGRNLVTMTARGLDGSGRATWFGLYNINSAGFFL	240


```
Db 1 MHKIVLAFALSLITIFALAVLLTSPGSS-----SQPHCPSSVSHRAPNPHPGSOQLF 56
Qy 61 ADLSREELTAVNRFLTORLGLVDAQAQPSDNCVFSVELQLPPKAAALHLDRGSPPP 120
Db 57 ADLSREELTAVNRFLTORLGLVDAQAQPSDNCVFSVELQLPPKAAALHLDRGSPPP 116
Qy 121 ABEALAIYFEGROPQPNVSELYVGLPHPSYMRDVTVEHNGRPLPHRRRPVFOEFLDD 180
Db 117 ABEALAIYFEGROPQPNVSELYVGLPHPSYMRDVTVEHNGRPLPHRRRPVFOEFLDD 176
Qy 181 QMIFNRELPOASGLHHCCEFYKHGRNLVTMTAPRGLOGSDRATFEGLYNISGAFPL 240
Db 177 RLKDELVEPKARIFLST--FVINGSTLAAVNAATPRGLSREKRTTIGLHINISGGLL 234
Qy 241 HHVGLLELNRKALDPARTIQKVFYQGRYVDSLAOLEAFEGAGLVNVLIPDNGTGSN 300
Db 235 HPVGLLELDRALDRAPHTVOOVFLGHYVADLGLEHFEKSGRLVVRPLPRNGAS 294
Qy 301 SLKSPVPPAPRLOFYPOGPRFSVSGSRVASSLMTFSGLGAFSGPRIFDVAFQGERLY 360
Db 295 SLRSRSPGLPPLQSPGSGSOYSVGNLVSSLSMSTFGHGVSGLRIFDVAFQGERLY 354
Qy 361 YEISLEALAIYGNRPAMATTRYVDGFGMGKTTPTLRGVDCPYLATVYDMHFLLESG 420
Db 355 YEVSVECVSITGADSPKMTLTRYLDSFGLNKGVLVRCVDCPYLATVYDMHFLLESG 414
Qy 421 APRTIRDAFCVEONOGDLRLRRHSDLYSHYEGGLAEVLYVNSMTLNYDYVMDTVPH 480
Db 415 AVOLLEAGVCFEEAOGDLRLRRHNYLQNHFYGLASALVYVSVSGYDYIMDFVLY 474
Qy 481 PGSAIEIRRYATGYISSATLFGATG--KGNQVSEHTLCTGVTHSHAFKVDLYAGLEW 538
Db 475 PNGAIEGRVHATGYINTALKGEGDLFGNREVERLGTVAHAFKLDLVAALKW 534
Qy 539 VVAEDVFPMAVPMSPSEHQLRLQVTRKLEKEEOAALFVGSATPRYLTLASNSKMG 598
Db 535 VVAEDVFPKVAAPNPMHMLORPOLITROYLKGEDLTAISGLSPRLYLASNQTNMG 594
Qy 599 HPRGRTIQLMSFAGEPLPONSSMARGFSWERQVLAVTQKKEEPSSSSVFNQNDPAVY 658
Db 595 HORG-----YOLVYTORKEESSOSSSIYHQNDIMTPTV 627
Qy 659 DESDFINNETIAGKDLVAVWTAGFLIHPAEDIPNTVYVNGNGVGLRPVNFEDDEPSY 718
Db 628 TRADFNNETLGEDLVAVWTASFLIHPAEDIPNTVTLGNRVGLLRPNFEDDEPSIF 687
Qy 719 SADSIFYRGDODAGACEVNPPLACLPOAACAPDLPAFSGCF 760
Db 688 SFGSYFEGQDAGLCSINPVACLPDLACVPLPFPFSGYHF 729
```

```
RESULT 4
US-10-005-530-58
; Sequence 58, Application US/10005530
; Publication No. US20030026795A1
; GENERAL INFORMATION:
; APPLICANT: Isaac, Barbara G.
; APPLICANT: Greenplate, John T.
; APPLICANT: Purcell, John P.
; APPLICANT: Romano, Charles P.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROLLING INSECTS
; FILE REFERENCE: 11899.0022.DVUS01 (MOBT-022--2)
; CURRENT APPLICATION NUMBER: US/10/005,530
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 09/063,733
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/044,504
; PRIOR FILING DATE: 1997-04-21
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58
; LENGTH: 617
; TYPE: PRF
```

```
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (74)..(79)
; OTHER INFORMATION: Xaa - Unknown
US-10-005-530-58
```

```
Query Match 2.6%; Score 106; DB 9; Length 617;
Best Local Similarity 22.8%; Pred. No. 1.4;
Matches 79; Conservative 40; Mismatches 141; Indels 86; Gaps 18;
```

```
Qy 63 LSREELTAVNRFLTORLG--PGLVDAQAQAR-----PSDNCVFSVELQLPPKAA 109
Db 18 LIREKAKELNTERLGEVPGIPRPNNEGFLGGYSHDNLPSDPLYSISK--PGGGG 73
Qy 110 LAHLDRGSPPARAEALAIYFEGROPQPNVSELYV-----PLPHPSYMRDVTVEHNGR 163
Db 74 XXXXXGCGPRK--VCIVAG-----VSGLYIAMLDDLKIPNLTYDIFESSRTGR 125
Qy 164 LPHRRPVLFQEXLIDIDOMIFNRELPOASGLHHCCEFYKHGRNLVTMTAPRGLOGCDR 223
Db 126 LYNHFTDAKHQYDIDGAMRY----PDPSMKRTFNLFRKTGMLI-----KYLIDGEN 175
Qy 224 ATTFGLYINISGAGFLHHVGELELVN-----HKALDPAMTQKVFYQGRYVDSLA 275
Db 176 TP--QLYNN--HFPAKGVSDPYVNVSVANGCTVPDDVYDSVEKIQAF--GYREKLA 227
Qy 276 QLEAFEGAGLVNVLIPDNGTGSWSLKSPVPPAPRLOFYPOGPRFSVSGSRVASSLW 335
Db 228 E--DFDKGDEMLVDMDMT--REYLKRGPRGGEAPKDYF-----FALQNMETON--- 273
Qy 336 TFSFGLGAFSGPRIFPVREGRLVYEISLQ-----EALAIYGN 376
Db 274 -----GTNLFDOAF--SESVDSFDPDNPTKEWYCIEGTS 309
```

```
RESULT 5
US-09-908-193-22
; Sequence 22, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: SHINKETS, RICHARD A.
; APPLICANT: ZERHUSEN, BRYAN
; APPLICANT: MALYANKAR, URIEL M.
; APPLICANT: PADIGARU, MURALIDHARA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 21402-062
; CURRENT APPLICATION NUMBER: US/09/908,193
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: 60/220,273
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: 60/221,650
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/221,233
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/220,912
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/218,875
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,870
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,901
; PRIOR FILING DATE: 2000-07-18
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 980
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-908-193-22
```



```

OY 174 OEYLDIDOMIFNRELPOASGLL-----HHCFYKHGRRLVMTTAPRGLOSGDRAT 225
DB 444 OYNLSISTSVGNVITPQAGGSLSLNGRDSKPHITDIDYGFELITSSAEVFTWANGDNKK 503
OY 226 WFLGYNISGAGFLLHVGLELVNHNKALDAPARMTQKVFYGRYDSIAQLEAO---FE 282
DB 504 RVLVLY--GGAG-ELHEFALPKHLRPRTVEGSYV--KIAQGSAMVWQEWMAARVRLR 558
OY 283 AGLVNVVILPONGGSGSLKSPVPPGAPPLQFY--PQGFPSVOGSRVASSL----- 335
DB 559 AGKLEIHLMRNDAYOHVLELPA---KOPIANYSPPSKETVIKGGGLLSAMITDND 614
OY 336 -----TFSPGLAFSGPRIFD--VRFOGERLYVEISLOEALAIIGNSPAA----- 379
DB 615 LHLGADVNTTPLEYISAPKRFDDIVENGOSIKSTRKIGLATAVHVQPPAISLPDLKR 674
OY 380 -----MTTRYVDGFGMGKRYTTPLTRGVDCPYLATYVDWHLFLESQAPKTIRDA 428
DB 675 LDMKYIDLPELSTFEXNDEGM-----TPLTN-----TY-----TNTREFTCP 712
OY 429 FCVEFQONGGLRRHNSDLSHYRGGLAEYVLVVRSMSTLLNYDYVMDTFHPGALIEIR 488
DB 713 TOLYADYVG-----YHGG-----SLIYRGHFT--ANGDESM--VFLNTSGG--VG 751
OY 489 FYATGYISSAFLEFGATGKYGNQVSEHTIGTVHTSAH-----FKYDLDAVAGLENMTWADM 544
DB 752 FANSWMLQOTELGSWTSGGRNMTYPRNITSLPHELSPGEPPYFTVVIDHNGOEAPGTDA 811
OY 545 VFVPAVWSPSEHOLRLQVTRKLEMEEOAFLVGSATPRYLYLASNH---SNKM---G 598
DB 812 IKFP-----RGILD-----VALSGHELSDLKMKMG 837
OY 599 HPRGRITMLSFAGEPLPQNSSMARGFSWERYOLAVTORKEEPPSSSVFNONDPAWPTV 658
DB 838 NLGEGQYDGLTRG--PLNEGAMTAER---OGYHL-----PSPPTSS--WKSSNP----- 879
OY 659 DESDEINNETIAGKDLVAMVATAGFLHIPHAEDIPNTVTVNGCVG-----FELRP 707
DB 880 -----IHEGLTAGIGIATYSFLDLPBGYDIPLSFPPNNSASARSGTSTRCOLFYNG 933
OY 708 YNF 710
DB 934 YQF 936

RESULT 10
US-09-738-626-4914
; Sequence 4914, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0

```

```

; SEQ ID NO 4914
; LENGTH: 581
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4914

Query Match      2.3% Score 95.5; DB 9; Length 581;
Best Local Similarity 22.8%; Pred. No. 11;
Matches 115; Conservative 50; Mismatches 188; Indels 151; Gaps 29;

OY 260 TIOKRYFOGRYDSIAQLEA--OFEAGLVNVY-----LIPDNGT--GGSMSSKSPVP 307
DB 164 TITKVIHDDGTEEGLAQLALQLEFKQVDYEDDLRETNASLPEDQITLAPRMGVATP 223
OY 308 PG---PAPPLQFYPOGPPRSVOGSRVASSLWTFSGLAFGSPRIF-----DYRF 354
DB 224 SDDKLAAPPLQSKRDGEMSPIV---IASP---ESGISAVATARAFAGAIEIRMDYDPDRF 276
OY 355 OGERLYVEISLOEALAIIGNSPAMTTRYVDGCGMGKTYTTPLRGVDCPYLATYVDNH 414
DB 277 NOEQKMYAGLADRLPLALDQ-----FGTG----- 302
OY 415 FLLESQAPKTIRDAFCVEONOG-----LPLRRHSDLYSH---YFGGLAE---TVLYVR 463
DB 303 ---EQLAEKIRGEVTELELPGGGGLVFPGRMVA--LIGHSPGALGALGEPAAEAVAR 358
OY 464 SKSTLLNTDYVDVTFHPSGALIEIRFYATGYISSAFLEGATGKYGNQVS--EHTIGTVH-- 520
DB 359 ATDMAQYQEFEDQVIDA-----FEVIAVYASBEA--GDQGNYSNEPVEDLVGYDAI 411
OY 521 THSAFXYVDLV--AGLEMKWVAEDVFPMAVPSPEHOLRLQVTRKL-----LEMEEO 574
DB 412 TEAGYAV--LDLQPGRAFNLQAKKY-----EELKRPNGALDPEMKIGPQEO 460
OY 575 AAFVGSATPRYLYLASNHNKMGHPRGYRIOMLSFAEPLPQNSSMARGFSWERYOLAV 634
DB 461 PMTRVGSADAELI-----NEVSQM-----LADL--AENDLPO-----KATVLAHQFLEM 503
OY 635 TORKEEPPSSSVFNONDPAWPTVPSDFINNETIAGDLVAMVATAGFLHIPHAEDINT 694
DB 504 LTRNDQ-----INTDLELAVLVHADGHN--AGDKYATW-----NM 538
OY 695 VTVNGVGFELRPYNEFDED--PSF 717
DB 539 LREGLSPIVTFMAKNFYDEDPPTF 562

RESULT 11
US-10-005-530-57
; Sequence 57, Application US/10005530
; Publication No. US20030026795A1
; GENERAL INFORMATION:
; APPLICANT: Isaac, Barbara G.
; APPLICANT: Greenplate, John T.
; APPLICANT: Purcell, John P.
; APPLICANT: Romano, Charles P.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROLLING INSECTS
; FILE REFERENCE: 11899, 0022, DVUS01 (MOBT-022--2)
; CURRENT APPLICATION NUMBER: US/10/005, 530
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 09/063, 733
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/044, 504
; PRIOR FILING DATE: 1997-04-21
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Polypeptide
; NAME/KEY: MISC_FEATURE

```

LOCATION: (74)..(79)
OTHER INFORMATION: Xaa = Unknown
US-10-005-530-57

Query Match 2.3%; Score 95; DB 9; Length 617;
Best Local Similarity 22.5%; Pred. No. 13;
Matches 78; Conservative 40; Mismatches 142; Indels 86; Gaps 18;

63 LSRELTAVMRELTORLG--PGLVDAQAAR-----PSDNCVFSEVQLPPKAAA 109
18 LIREKAKELNLTRELGVPPIPRNEGRLGGVSHNDLPDLYSSI-----GGSGG 73
110 LAHLDGSPPARREALATYFFGHQPOPNTSELVYG-----PLPHSYARDVVERHGG 163
74 XXXXXGGGPPRR--VCYIAGAG-----VSGLYAMILDCLKIPNLTYYDFESSSRGG 125
164 LPYHRRPVLFQELIDOMIFNRELPOASGLHHCCEFYKHGRNLTVMTPAPRGLQSDR 223
126 LTHHTTDAKHXYDIGAMRY-----PDIPSMKRTFNLKRTGMPLI-----KYYLDGEN 175
224 ATWFGLYNISGAGFLHVGLELVN-----HKALDPARKWTQKVYFGRYDLSA 275
176 TP--QLYNN-----HFFAKVSDPYMVSANGTVPDDVDSVGEKLOQAF--GYREKELA 227
276 QLEAOFEAGLVNVLIPDNGTGSNLSKSPVPGAPPLQFYPOGRFSVQSRVASSLM 335
228 E--DDEKGFDELMDVDMTT--REYLKRGKPGKAPKYDF-----FALQMETON-- 273
336 TFSFGAGFSPPRIFDVREFGERLYEISLQ-----EALAIYGSN 376
274 -----TGTNLFDOAF--SESVIDSPEDNTRKEPMTCIEGTS 309

RESULT 12

US-09-912-020-364

Sequence 364, Application US/09912020

Patent No. US20020045592A1

GENERAL INFORMATION:

APPLICANT: zyskind, Judith

APPLICANT: Ohlsen, Karl L.

APPLICANT: Trawick, John

APPLICANT: Forsyth, R. Allyn

APPLICANT: Froelich, Jamie M.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN

FILE REFERENCE: ELITRA.001DVI

CURRENT APPLICATION NUMBER: US/09/912,020

CURRENT FILING DATE: 2001-07-23

PRIOR APPLICATION NUMBER: 09/492,709

PRIOR FILING DATE: 2000-01-27

PRIOR APPLICATION NUMBER: 60/117,405

PRIOR FILING DATE: 1999-01-27

NUMBER OF SEQ ID NOS: 485

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 364

LENGTH: 878

TYPE: PRF

ORGANISM: E. coli

US-09-912-020-364

Query Match 2.3%; Score 93.5; DB 10; Length 878;
Best Local Similarity 20.2%; Pred. No. 30;
Matches 108; Conservative 61; Mismatches 190; Indels 175; Gaps 27;

176 YLDIDOMIFNRELPOA-----SGLLHHCCEFYKHGRNLTVMTP-- 212
148 HLDVGGQRLNLTIPQAFMSNRARGYIPPELMDPGINAGLLN---YNFSGNSVONRIGN 203
213 --TAPRLOSG-----DRATWFGLYN---ISGAGFLHVGLELVNHNKALDP- 256

204 SHKATINLOSGLNIGAMRLDNTTW---SYNSDRSSGSKNKKQHINTWL---ERDIIPL 257

257 -ARMTIQKVYQGRYDSI---AQLEAOFEAGLVNVLIPDNGT----- 297

258 RSRLLTGDDYTGODIPDGINFRAQLASDNN-----MLPDSORGFAPVINGIANGTAQ 310

298 -----GSMLSKSPVPPGAPPLQFYPOGRFSVQ-----GS-----R 329

311 VTIKONGYDIYNSTVPPGPFITINDIYAAGSGDLQVITKEADSGTIFTPVSSVPLLR 370

330 VASLMTFSFGLCAFSGPRIFDVRFQGERLYEISLOEALATYGSNPAAMTRVYDGEF 389

371 EGHTRISTNAGEERSGNAOQEKTRFFOSILLH--GLPAGTITYGGO---LADRYAFNF 425

390 GMGKYTTPPLTRGVDCPYLATYVDMHFLLESQAKTIRDAFCVEQONOGPLRRHNSDLS 449

426 GICKNMGALG-----ALSDM-----TQANSTLPDD---SOHDGQSVR---FLYN 464

450 HYFGGLAEVLYVRSMTLLNIDYVMDTVPH-----PSGAIERFATGYISSAPL 500

465 KSLNESGTNIQLVGYRISTSGYFNADTTYSRMNGINIEFQDVCVIOKPKFTDYNDLA-- 522

501 FGATGKYGNQVSEHTLG--TVHTSHAH-----FKYDLVAGLENNVMAEDMV 545

523 YNKRGLQLTLVTOO--LGRFTSLYLSGSHQYWGTSNVDEQFOABLNTA-----FED-- 572

546 FVPMVAVPMSPEHOLQRLQVTRKLEMEBQAFLVGSATPRYLILASHNSKKMGH 599

573 -----INWLTLSYL-----TKNAMQKGRDMLANVINIPSHWLRSDSKSKQMH 616

RESULT 13

US-09-734-569-172

Sequence 172, Application US/09734569

Patent No. US20020064816A1

GENERAL INFORMATION:

APPLICANT: Lersch, Jens

APPLICANT: Renz, Andreas

APPLICANT: Ehrhardt, Thomas

APPLICANT: Reindl, Andreas

APPLICANT: Clrpus, Petra

APPLICANT: Bischoff, Friedrich

APPLICANT: Frank, Markus

APPLICANT: Freund, Annette

APPLICANT: Duwenig, Elke

APPLICANT: Schmidt, Ralf-Michael

APPLICANT: Reski, Ralf

TITLE OF INVENTION: Mose genes from Physcomitrella patens encoding proteins invc

FILE REFERENCE: BASF-NAE-1332-99-US

CURRENT APPLICATION NUMBER: US/09/734,569

CURRENT FILING DATE: 2001-05-24

PRIOR APPLICATION NUMBER: US 60/171,101

PRIOR FILING DATE: 1999-12-16

NUMBER OF SEQ ID NOS: 181

SOFTWARE: PatentIn Ver. 2.1/WordPerfect 6.1

SEQ ID NO 172

LENGTH: 480

TYPE: PRF

ORGANISM: Physcomitrella patens

US-09-734-569-172

Query Match 2.3%; Score 93; DB 10; Length 480;
Best Local Similarity 22.9%; Pred. No. 13;
Matches 91; Conservative 55; Mismatches 124; Indels 128; Gaps 22;

1 MNQTIYLVLLIAYI-----TFALYCV--LLVGRGDDGEGEPQLPHCSYS 45
168 LNDLFCVIGILLALDPLGNSPNWMTWTFALATVAVVLGLD-----MAYCPE-S 218
46 PSAQPMTHPQSOQLFADLSRELTAVMRELTORLGVPGLVDAQAARPSDNCVFSEVQLDLP 105
219 PR--WLTKKK-----TAEATVAVRL-----WGAKKVESNADLAASSVETVK----- 260

QY 106 KAALAHLDGRSPPPAREALAIFFGROPQPNVSELYVGLPHPSYMDVVERHGGPLP 165
Db 261 -----GDTODASNGELFGKR-----YRKYVTV---GMAL- 286
QY 166 YHRRPVLEOYLIDQMI-FNRELPOASGLHHCCFKHGRNLTMTTARGL--QSGD 222
Db 287 -----FLEOGAGINAVYFSTGVFSAGITNDVAASALVAGNAVAGTVAAGMDKGR 341
QY 223 RAYFGLYXNSGAGFLHVGLELLVNHKALDPARTTQKVFYQGRYDSLAOLEAOFE 282
Db 342 KSLMG-----SFGAGMSLMLVLSLMSWPLAP-----YSGTIAVL----- 378
QY 283 AGLVNVLPDNGSGMSLSKSPVPPGAPAPLQFYPO--GPRFSVQSGRVASSL---WTF 337
Db 379 -GYSTYL-----SFSIGA-----GPVREL-LLELIFGAR--IRAKAVALSLGVMTC 422
QY 338 SFGIAGSFPRIFFDVRFOGERLYVEISLQELALAIYGCN 375
Db 423 NFMIGLEFLNVOYKFGVSTVYLFPSAVCAALIVVGN 460

RESULT 14

US-09-925-301-844
; Sequence 844, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925.301
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 844
; LENGTH: 601
; TYPE: PR
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (36)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (64)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (103)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (106)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (152)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (358)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (383)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-844

Query Match 2.38; Score 93; DB 10; Length 601;
Best Local Similarity 22.3%; Pred. No. 19;
Matches 78; Conservative 32; Mismatches 112; Indels 128; Gaps 19;

QY 31 DGGSPQLPHC-----PSVSPSA-QPWTHP-----GOSLFDLSREE 67
Db 218 DTGTEALTPHLMRNLHATSRKSTYRPSMERMPELSPFEDVAGTEMSQSDSGVLDSDS 277

QY 68 LTAVMRELTORLGP--GLYDAQAARP-----SDNCVSEYELQPPKAAALAHLDGRSPPA 121
Db 278 QVS--SGPCSQRSSPDGGLKGALEGPPKRRPGSSPLNAPCEGPP-----GSEPPR 326
QY 122 REALAIVFGRGROPQPNVSELYVGLPHPSYMDVTVR--HGGL-PYHR--RPVLF--- 173
Db 327 RPPAPHDGDKELPREOPPLPPII---GTRSOXTRDGTPEGRTIRSHRGRPPVQGTGX 383
QY 174 QEYLDI-----DOMIFNRE-----LPQASGLHHCCFKHGRNLTMT 211
Db 384 DKDSDLRLVVGDSLKAKEKELTASVTEALPYSRDWMELLPAAASA-----EPOSKNDSC 437
QY 212 TTAARGDSGRATWFGLYNISCAGFLHVGLELLVNHKALDPARTTQKVFYQGRY 271
Db 438 HCVEPSSSGR-----LYPEVFF 456
QY 272 DSLAOLEAOFEAGLVNVLPDNG--TGSMSL-----KSP--VPPGAPP 313
Db 457 GSAPSSSQISGAMDSDLPHNSGCFRPTSLPHYNSQPLYLPFGAPP 506

RESULT 15

US-08-954-701A-19
; Sequence 19, Application US/08954701A
; Publication No. US20030032085A1
; GENERAL INFORMATION:
; APPLICANT: SCOTT, MATHEW P
; APPLICANT: GOODRICH, LISA V
; APPLICANT: JOHNSON, RONALD L
; TITLE OF INVENTION: Patched Genes and their Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley, Hoag & Eliot
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/954,701A
; FILING DATE: 20-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36709
; REFERENCE/DOCKET NUMBER: SUV-003.08
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1447 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-954-701A-19

Query Match 2.38; Score 93; DB 1; Length 1447;
Best Local Similarity 25.5%; Pred. No. 71;
Matches 49; Conservative 28; Mismatches 67; Indels 48; Gaps 10;

QY 6 ILVLLIATVITIFLVVLL--VORGSDGGEPSQLPHCPSPVSPAQ-----WTH 53
Db 1156 LAITLIGLVLLGLVLLVLSFPGPIEVSBNAGLNLPTSPSPPPSVVRFAMPQETH 1215
QY 54 PGOSQLFADLSREE---LTAVMRELTORLGP-----LYDAQAARPSDNCVSEYELQ 103

```
Db 1216 SGSDSDSEYSSQTVSGLSSELRLHYEAQGAGCPAHQVIYEA-----TENPVFAHSTV 1270
OY 104 PPKA-----ALAHLDRGSPPPAREALATVFGROPQPN-VSELYVGPLPHPSYMD 154
Db 1271 HPESRHHPPSNPROQPHLDSGSLPPGRQ-----GQPRRDPPREGLWPLYRP--RD 1321
OY 155 ---VVERHGR 163
Db 1322 AFEISTEGHGP 1333
```

Search completed: May 20, 2003, 12:02:20
Job time : 127.567 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:09:54 ; Search time 29.4394 Seconds
(without alignments)
2491.578 Million cell updates/sec

Title: US-10-081-408-2
Perfect score: 4095

Sequence: 1 MNOKTILVLLILAVITTFAL.....QAAACAPDLPAFSHGFSHN 763

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4095	100.0	763	2 JC5234	amine oxidase (cop
2	3334	81.4	762	2 A54411	amine oxidase (cop
3	1400.5	34.2	746	2 S34656	amine oxidase (cop
4	1394	34.0	751	2 A54053	amine oxidase (cop
5	413	10.1	755	2 B41836	amine oxidase (fla
6	397.5	9.7	757	2 E64889	amine oxidase (cop
7	377	9.2	684	2 A56102	amine oxidase (cop
8	368.5	9.0	660	2 G90330	amine oxidase (cop
9	363	8.9	648	2 A48646	amine oxidase (cop
10	357.5	8.7	756	2 T48139	amine oxidase (cop
11	353.5	8.6	650	2 G71412	copper amine oxida
12	340	8.3	638	2 JC2139	probable amine oxi
13	332	8.1	674	2 C48130	phenylethylamine o
14	327	8.0	687	2 T47403	amine oxidase (cop
15	320	7.8	674	2 C44239	amine oxidase-like
16	302.5	7.4	759	2 E64854	amine oxidase (cop
17	300	7.3	660	2 A82234	probable copper am
18	295	7.2	692	2 S04963	copper amine oxida
19	247	6.0	794	2 T39171	amine oxidase (cop
20	246.5	6.0	300	2 T48138	probable peroxisom
21	229	5.6	587	2 S21139	copper amine oxida
22	214.5	5.2	670	2 S71320	amine oxidase (cop
23	126	3.1	862	2 T46289	amine oxidase (cop
24	125.5	3.1	460	2 T48137	hypothetical prote
25	125	3.1	757	2 C84120	copper amine oxida
26	111	2.7	1322	2 D82685	subtilisin-type pr
27	107	2.6	435	2 A42672	phosphoribosylform
28	106.5	2.6	896	2 F96523	choline kinase (EC
29	106	2.6	3938	2 T42761	hypothetical prote
					Bassoon protein -

30	103	2.5	2224	1 KFU05	coagulation factor
31	102.5	2.5	591	2 C48364	cytochrome-c oxida
32	101.5	2.5	546	2 B75573	conserved hypothet
33	101.5	2.5	848	2 C65083	hypothetical prote
34	101	2.5	626	2 T35669	hypothetical ATP/G
35	100.5	2.5	597	2 C67381	sensor histidine k
36	100.5	2.5	654	2 S76870	hypothetical prote
37	100.5	2.5	1257	1 A41060	hypothetical prote
38	100.5	2.5	1544	2 E59431	neural cell adhesi
39	100	2.4	919	2 S42842	phosphoinositide-b
40	99.5	2.4	579	2 E83144	T16G12.2 protein -
41	99.5	2.4	2352	2 C83229	hypothetical prote
42	99	2.4	602	2 T35782	probable non-ribos
43	99	2.4	602	2 T35782	probable secreted
44	99	2.4	915	2 T03589	probable aspartate
45	98.5	2.4	2055	2 T00093	hypothetical prote
			881	2 S25445	nitrate reductase

ALIGNMENTS

RESULT 1

JC5234
amine oxidase (copper-containing) (EC 1.4.3.6) AOC2 precursor - human

C:Species: Homo sapiens (man)
C:Date: 13-Mar-1997 #sequence_revision 13-Mar-1997 #text_change 18-Feb-2000

C:Accession: JC5234
R:Zhang, X.; McIntire, W. S.

Gene 179, 279-286, 1996

A:Title: Cloning and sequencing of a copper-containing, topa quinone-containing mon

A:Reference number: JC5234; M01D:97128319; PMID:8972912

A:Contents: Placenta

A:Accession: JC5234

A:Status: Preliminary

A:Molecule type: mRNA

A:Residues: 1-763 <ZHA>

A:Cross-references: GB:U039447; NID:g1399031; PIDN:AAC50919.1; PID:g1399032

C:Comment: This enzyme catalyzes the oxidation of primary amines to the correspondi

Cu(II) and 1 molecule of covalently-bound topa quinone. It also can scavenge cirru

C:Genetics:

A:Gene: GDB:AOC2; DAO2

A:Cross-references: GDB:4562632

A:Map position: 17q21-17q21

C:Superfamily: amiloride-binding protein

C:Keywords: copper; metalloprotein; oxidoreductase; quinoprotein; topaquinone

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-763/Product: amine oxidase (copper-containing) #status predicted <MAT>

F:444-520-522/Binding site: copper (His) #status predicted

F:471/Modified site: topaquinone (Tyr) #status predicted

C:Genetics:

Query Match 100.0%; Score 4095; DB 2; Length 763;
Best Local Similarity 100.0%; Pred. No. 1.5e-316;
Matches 763; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 301 SLKSVPPGAPAPLQFYOGGPRFSVQGSRVASLSLMTFSFGGATSGSPRIEDVRRQGEELY 360
 Db 301 SLKSVPPGAPAPLQFYOGGPRFSVQGSRVASLSLMTFSFGGATSGSPRIEDVRRQGEELY 360
 Oy 361 YEISLQELALATYGGSNPAAMTRYVDGFGMGKYYTPTLRGVDCEPYLATYVDHMFLESQ 420
 Db 361 YEISLQELALATYGGSNPAAMTRYVDGFGMGKYYTPTLRGVDCEPYLATYVDHMFLESQ 420
 Oy 421 APKTRDAFCVFEQNGQLPLRRHSDLYSHYFGGLAETVLVVRSMSTLLNDYWDVDFH 480
 Db 421 APKTRDAFCVFEQNGQLPLRRHSDLYSHYFGGLAETVLVVRSMSTLLNDYWDVDFH 480
 Oy 481 PSGLAEIRFVYATGYISSAFELGATGKXGNOVSEHTLGVHSHAFKFDLDVAGLENNYV 540
 Db 481 PSGLAEIRFVYATGYISSAFELGATGKXGNOVSEHTLGVHSHAFKFDLDVAGLENNYV 540
 Oy 541 AEDWVPMAYPWPSEHQLQRLQYTRKLLMEEDQAAFLVGSATPRYLYLASNHSKMGHP 600
 Db 541 AEDWVPMAYPWPSEHQLQRLQYTRKLLMEEDQAAFLVGSATPRYLYLASNHSKMGHP 600
 Oy 601 RGYRQLMLSPFGCEPLPONSNAARGESWERYOLAVYORKKEEPPSSSVFNQNDPMAPYDF 660
 Db 601 RGYRQLMLSPFGCEPLPONSNAARGESWERYOLAVYORKKEEPPSSSVFNQNDPMAPYDF 660
 Oy 661 SDFINNETIACKDLVAMYTAGEFLHPHAEEDIPNVTYVNGNGEFLRPVNFEDDEDPSEYSA 720
 Db 661 SDFINNETIACKDLVAMYTAGEFLHPHAEEDIPNVTYVNGNGEFLRPVNFEDDEDPSEYSA 720
 Oy 721 DSIYFRGDQDAGACEVNPDLACLPLQAAACAPDLPAFSGHGFSGHN 763
 Db 721 DSIYFRGDQDAGACEVNPDLACLPLQAAACAPDLPAFSGHGFSGHN 763
 RESULT 2
 A54411
 amine oxidase (copper-containing) (EC 1.4.3.6), serum, precursor - bovine
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 09-Sep-1994 #sequence,revision 06-Feb-1995 #text_change 18-Feb-2000
 C:Accession: A54411; B38081; A48242; S65408
 R:Mu, D.; Medzhardsky, K.F.; Adams, G.W.; Mayer, P.; Hines, W.M.; Burlingame, A.L.; Smith, J.P.
 A:Title: Primary structures for a mammalian cellular and serum copper amine oxidase.
 A:Reference number: A54411; MUID:94193686; PMID:8144587
 A:Accession: A54411
 A:Molecule type: mRNA
 A:Residues: 1-762 <MUA>
 A:Cross-references: GB:S65953; NID:9546215; PIDN:AAH30397.1; PID:9546216
 R:Mu, D.; Jones, S.M.; Smith, A.J.; Brown, D.E.; Dooley, D.M.; Kilman, J.P.
 J. Biol. Chem. 267, 7979-7982, 1992
 A:Title: Tyrosine codon corresponds to topa quinone at the active site of copper amine oxidase.
 A:Reference number: A38081; MUID:92235001; PMID:1569055
 A:Accession: B38081
 A:Molecule type: Protein
 A:Residues: 463-465, 'D', 467-473, 'X', 475-485 <MU2>
 R:Jones, S.M.; Mu, D.; Wemmer, D.; Smith, A.J.; Kaur, S.; Matthy, D.; Burlingame, A.L.; Science 248, 981-987, 1990
 A:Title: A new redox cofactor in eukaryotic enzymes: 6-hydroxydopa at the active site of amine oxidase.
 A:Reference number: A48242; MUID:90260648; PMID:2111581
 A:Accession: A48242
 A:Status: Preliminary
 A:Molecule type: Protein
 A:Residues: 468-472 <JAN>
 Ride Blase, D.; Apostolli, E.; de Matteis, G.; Mondovl, B.; Morpurgo, L.
 Eur. J. Biochem. 237, 93-99, 1996
 A:Title: Half-of-the-sites reactivity of bovine serum amine oxidase. Reactivity and chemical mechanism of the active site.
 A:Reference number: S65408; MUID:96203913; PMID:862089
 A:Accession: S65408
 A:Molecule type: Protein
 A:Residues: 463-469, 'X', 471-487 <DEB>
 C:Superfamily: amiloride-binding protein
 C:Keywords: copper; glycoprotein; oxidoreductase; quinoprotein; topaquinone
 F1-16/Domain: signal sequence #status predicted <SIG>

F:117-762/Product: amine oxidase (copper-containing) #status predicted <MAT>
F:136,231,665/Binding site: carboxylate (Asn) (covalent) #status predicted
F:470/Modified site: topaquinone (Tyr) #status experimental

Query Match	81.4%	Score 3334	DB 2	Length 762
Best Local Similarly	80.68%	Pred. No. 3.6e256		
Matches 610; Conservative	61;	Mismatches 76;	Indels 10;	Gaps 2

QY	14	VITTAALVCVLLVGC--GGDGG-----PSQLPHCPSPSPSAQPTTHGQSGFLD	63
Db	3	IFIFSLTTLVLMGEEGGVSGSEGVKQCHPSLRPCPCSPSPSPPTHPDSQFLD	62
QY	64	SREELTAVMRELTLQGLGDLVDAQAARPSDNCVSEVELQLPKKAALAHLDGRSPARE	123
Db	63	SREELTVMSTLQGLGDLVDAQAARPSDNCVSEVELQLPKKAALAHLDGRSPARE	122
QY	124	ALATYFSGROQPNVSELVGBPLPHPSYMDVYVERHGGPLPHRRPVLFQBYLIDOMI	183
Db	123	ALATYFSGGQPNVTELVGPPLPQPSYMRVYVERHGGPLPYHRRPVLLREYLDIDOMI	182
QY	184	FNRBELPQASGLLHHCFCFKHGRNLYTMTDAPRGISQDBRATMGLYINISGAGFELHHV	243
Db	183	FNRBELPQAGVLLHHCFCSTKGGKQLTNMSAPRGQSDNRSTWETLYNITKGGYLLHPV	242
QY	244	GLELLVNHKALDPAKMTLOKVFYQGRYDLSLAOLEACEAGLVNVLIPDNGTGGSMELK	303
Db	243	GLELLVNHKALDPAKMTYQKVFQGRYENLAQLEQGEAGQVNVVYIPDGTGGFMSLK	302
QY	304	SPVPPGAPPLLOFTYQGRFVSQGRVASSLMTSESGLCATSGPRIDFVRQGERLYEI	363
Db	303	SOVPPGAPPLLOFHQGRFVSQGRVASSLMTSESGLCATSGPRFDFVRQGERLAYEI	362
QY	364	SLQELALTYGGNSPAAMTTRVYDGGFCMGKXTTTLPLRVDQPYLATYVDMHFLESOAPK	423
Db	363	SLQELGAVYGGTAPAAMLTTRVYDGGFCMGYFATPLIRKVDQPYLATYVDMHFVVESSQPK	422
QY	424	TIRDAFCVFEQNOGRLPRRHSDLYSHYFGLGAETVLVVRMSLTLANDYVMDYFHRSG	483
Db	423	TLHDACVCFEQQNGKRLPRRHSDPLSHYFGVAGQVVLVFRSYSTLANDYVMDVYFYRNG	482
QY	484	ALIEIRFVATGYISSAFLEFGATGKYGNQVSEHTLGTVHTSHAHFKVYDLVDAGLENNVMAED	543
Db	483	ALIEVLAHTGTYSFAFLGGAARRYGNQVSEHTLGTGVHTSHAHKYVDLDVGGLENNVMAED	542
QY	544	MYFVPMVMPWSEHLOLOLYTRKLLMEDEQAFLVGSATPRYLYLANSHSKMKHPRGX	603
Db	543	MAFVPTALPMWSEHIOIRLYTRKQLETEEQAALFGLGASPRYLYLASKOSKMKHPRGX	602
QY	604	RLOMSFAGEPLPQOSSNARGFSMERYOLATYORKEEPPSSSYVENQNDMPAPTYDFSGF	663
Db	603	RLOQYSPFAGGPPQOSSPMERAFSGRQYOLATYORKETEPSSSYVENQNDPMTPTYDFSGF	662
QY	664	INNETIACKDLVAMVTAGFLHIPHAEDIPNTYVNGVGVFLRPYNFPEDEDPSPYSADSI	723
Db	663	INNETIACKDLVAMVTAGFLHIPHAEDIPNTYVNGVGVFLRPYNFPEDEPSMDSADSI	722
QY	724	YFRGQODAGACEVNPPLACLPOAAACAPDLPAFSHSGF 760	
Db	723	YFRGQODAGCEINPLACLPOAATCAPVPFSSHGY 759	

RESULT 3
S34656
amine oxidase (copper-containing) (EC 1.4.3.6), kidney, precursor - rat
N:Alternate names: amiloride-binding protein, long form
N:Contains: amiloride-binding protein, short form
C:Species: Rattus norvegicus (Norway rat)
C:Date: 13-Jan-1995 #sequence revision 13-Jan-1995 #text_change 18-Feb-2000
C:Accession: S36847; S36648; I51904; S34656; S34657
R:Lingnegalla, E.; Renard, S.; Volley, N.; Waldmann, R.; Chassande, O.; Lardunski, M.
Eur. J. Biochem. 216, 679-687, 1993
A:Title: Molecular cloning and functional expression of different molecular forms of
A:Reference number: S36847; MUID:93387321; PMID:8375402

A:Accession: S36847
 A:Molecule type: mRNA
 A:Residues: 1-746 <LIN2>
 A:Cross-references: EMBL:X73911; NID:g395064; PIDN:CAA52116.1; PID:g395065
 A:Accession: S36848
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 529-746 <LIN2>
 A:Cross-references: EMBL:X73912; NID:g395066; PIDN:CAA52117.1; PID:g395067
 A:Title: Isolation of a rat amiloride-binding protein cDNA clone: tissue distribution an
 A:Reference number: 151904; MUID:94295632; PMID:8023885
 A:Accession: 151904
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 530-586, 'r', 588-632 <VER>
 A:Cross-references: GB:S70383; NID:g547214; PIDN:AA831157.1; PID:g547215
 A:Note: In Genbank entry S70383, release 116.0, the source is designated as Rattus sp.
 C:Keywords: alternative initiators; oxidoreductase
 C:Superfamily: amiloride-binding protein
 F:1-746/Product: amine oxidase (copper-containing), kidney, precursor #status predicted <SHRT>
 F:529-746/Product: amiloride-binding protein short form #status predicted <SHRT>

Query Match 34.2%; Score 1400.5; DB 2; Length 746;
 Best Local Similarity 39.1%; Pred. No. 9.1e-103;
 Matches 300; Conservative 131; Mismatches 280; Indels 57; Gaps 14;

18 FALVCVLLVGRGSGDEPQLPHCPSPVSAQPFTHPGOSLPFADLSREELAVMFLNQ 77
 1 FVLLVLLVGRGSGDEPQLPHCPSPVSAQPFTHPGOSLPFADLSREELAVMFLNQ 77
 5 FGMANVILVQLQYDTASAVRTP-----DARVADLSPOEIKAVHSFLAN 50
 78 RLPGGLVDAQAARSDNCVFSVELQLPKAAHLADKSPPARALAVFEGROPQN 137
 1 RLLVLLVGRGSGDEPQLPHCPSPVSAQPFTHPGOSLPFADLSREELAVMFLNQ 77
 51 REELGQPSKEPTLANKNSFLLEMLPKKKHVKFLDEGKGFNRRAAVIFFGADYPPN 110
 138 VSELVYGPLPHPSYMDVVERHGSLPYHRRVPLFOEYLDIDQMIFNRELPOASGLIH 197
 1 VSELVYGPLPHPSYMDVVERHGSLPYHRRVPLFOEYLDIDQMIFNRELPOASGLIH 197
 111 VTEFVAGPLPRPYTRIALS-PRGHHLSWSSRPISTAEY---DLVTH-TLKRAVPLHQ 164
 198 -----CCEYKHKGRNLYMTTAPRGLOSGDRATWFGLYYNSGAFELHNGELLVN 250
 165 FLDDTGFSEFGDRCCLFTVDVAPRGVASGGRSMFVYQRYVE--GYELHPRGLEILD 222
 251 HKAIDPARWTKQVFTQGRYDSLAOLEAFOEAGLVNVVLLPD---NGTGSNSLSKSPV- 306
 1 HKAIDPARWTKQVFTQGRYDSLAOLEAFOEAGLVNVVLLPD---NGTGSNSLSKSPV- 306
 223 HGSTDQVDRVEQLWNGKFLNPNBELARKYAVGEVDVLEDPPLNGTE----KPLPF 277
 307 -----PPGAPRPLQFTYQGRFVSQGSRVASLSLMTFSEFGAGSPGRIPTDYRQ 355
 278 SSTKPRGEHTTPVNAVGPVAVGSGPRYKLEGMTVLXGGSFYRLRSSGLDIFNVLEFG 337
 356 GERLYVEISLOEALAIYGSNSPAAATTRYVDGFGMGKTYTTPTRGVDCPYLATYVDMHF 415
 338 GERVAEVSQAEVALYGGHTPRAGMOTKYIDVGMGASVTHELAPGDCETTFIDAFH 397
 416 LLESQAPKTIKRAFCVFEONOGIPLARRHSDLYS---HFGGLAEFLVLVVRSNTLLND 472
 1 LLESQAPKTIKRAFCVFEONOGIPLARRHSDLYS---HFGGLAEFLVLVVRSNTLLND 472
 398 YYDSQDPVHYPHALCLFEKPTGVPRLRRHNSNKGGFNFAGLKGVLVLRITSTSYND 457
 473 YVVDVTFHPSGAIEIRFYATGYSSAFLGATGKYGNOVSEHRLGLVYTHSAFXYDDV 532
 458 YLMDYLFYSNGVEMAKMHAATGYHATFYREGLRHGTRLQTLGLGNIHLVLRVDMDV 517
 533 AGLENVMAEDVAVFVMAVPMSPERHQLQRLVTRKLEMEQQAFLVGSATPRYTLASN 592
 518 AGTKNSFQRLTKMLKLTNPWSPSHLVQPTLEQTOYSGEHAQAFRGCQRLPYTLFESSP 577
 593 HSKMKHPRGRTQMLSPAGEPLPONSNAKGSWERYQLAVTQKKEEPSSSSVENQND 652
 578 OKKCMGHRSYRLQIHSMAEQVLPFGMOEERATVTAARYPLAVTKYSESEYSSSLYQND 637
 653 PMAPTVDFSDFI-NNEFTIAGKDLVAVWVTAAGFLIIPHAEDIPNVTVYTGNGNGFLLRPYNF 711

DB 638 PMDPVVEEELFKNNENIDEDLVAVWVTAAGFLIIPHSEVDPNATGNSVGLFLRPYNF 697
 712 DEDPSFYSADSIYFRGDQACAEVNPRLACLPOAAACADLPAPFSGHG 759
 DB 698 PEDPSLASRDYIV-WPQKGLNRYQ--RWIPEDRCIVS-PPFSYNG 741

RESULT 4
 A54053
 amine oxidase (copper-containing) (EC 1.4.3.6) ABP1 precursor - human
 N:Alternate names: amiloride-binding protein; diamine oxidase; histaminase
 C:Species: Homo sapiens (man)
 C:Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 18-Feb-2000
 C:Accession: A54053; A54410; I38708; A38276; S42495
 R:Chassande, O.; Renard, S.; Barbry, P.; Lazdunski, M.
 J. Biol. Chem. 269, 14484-14489, 1994
 A:Title: The human gene for diamine oxidase, an amiloride binding protein. Molecula
 A:Reference number: A54053; MUID:94237856; PMID:8182053
 A:Accession: A54053
 A:Molecule type: DNA
 A:Residues: 1-751 <CHAS>
 A:Cross-references: EMBL:X78212; NID:g463242; PIDN:CAA55046.1; PID:g463243
 R:Novotny, W.F.; Chassande, O.; Baker, M.; Lazdunski, M.; Barbry, P.
 J. Biol. Chem. 269, 9921-9925, 1994
 A:Title: Diamine oxidase is the amiloride-binding protein and is inhibited by amilo
 A:Reference number: A54410; MUID:94193685; PMID:8144586
 A:Accession: A54410
 A:Molecule type: protein
 A:Residues: 20-27, 'A', 29-39 <NOV>
 A:Cross-references: PIDN:AA830395.1; PID:g546213
 A:Experimental source: placenta
 A:Note: sequence extracted from NCBI backbone (NCBIP:146046)
 R:Zhang, X.; Kim, J.; McIntire, W.S.
 Biochem. Genet. 33, 261-268, 1995
 A:Title: CDNA sequences of variant forms of human placenta diamine oxidase.
 A:Reference number: I38708; MUID:96113540; PMID:8595053
 A:Accession: I38708
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-331, 'F', 333-751 <RES>
 A:Cross-references: EMBL:U11862; NID:g533535; PIDN:AA50270.1; PID:g533536
 R:Barbry, P.; Champe, M.; Chassande, O.; Munemitsu, S.; Champigny, G.; Lingueglia, P.
 Proc. Natl. Acad. Sci. U.S.A. 87, 7347-7351, 1990
 A:Title: Human kidney amiloride-binding protein: cDNA structure and functional expr
 A:Reference number: A38276; MUID:91017502; PMID:2217167
 A:Accession: A38276
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-265, 'RCLGARGMTAQRSPSPPPQARDFPPQHPCEPRLGAPRPSIQAGGRCALRLELCPV',
 A:Cross-references: GB:M55602; GB:M56335; NID:g387655; PIDN:AA58358.1; PID:g177960
 A:Note: this sequence has been corrected in reference A54053
 C:Genetics:
 A:Gene: GDB:ABP1
 A:Cross-references: GDB:127105; OMIM:104610
 A:Map position: 7q31-q32
 A:Introns: 524/1; 619/2; 663/3
 C:Function:
 A:Description: oxidation of putrescine and histamine
 C:Superfamily: amiloride-binding protein
 C:Keywords: homodimer; oxidoreductase; quinoprotein; topaquinine
 F:1-13/Domain: signal sequence #status predicted <Sig>
 F:461/Modified site: topaquinine (Tyr) #status predicted

Query Match 34.0%; Score 1394; DB 2; Length 751;
 Best Local Similarity 39.8%; Pred. No. 3e-102;
 Matches 293; Conservative 124; Mismatches 281; Indels 38; Gaps 11;

50 PMTHPGOSLPADLSREELAVMFLNQRLGPGVDAQAARSDNCVFSVELQLPKAA 109
 DB 23 PGLTPKAGVSDLSNDELKAVHSFLMSKKEELQPSSTTMAKNVFLIEMLPKKYHV 82
 110 LAHLDSGPPPARALAVIFFGROPQPNVSELVYGPLPHPSYMDVVERHGSLPYHRR 169

```

Db      83 LRLDGERHPRVEARAVIFFGDEHPNNTFEAVGCLPBGCVARALS -PRPGVSSMASR 141
      170 PVLFOEYLDIDOKIFNRELPOASGLHH-----CCFKHGRNRLVMTTARGLOSD 222
      142 PLSTASY-----ALVHTLOEAKRPLHQFLLNTGFSFDCDHRCLATFDVARGVASGQ 196
      223 RMTWFLYNYISGAGFLLHVGLLELVNHNKALDPAWITOKVYOGRYDLSLAQLEAOFE 282
      197 RSMWLLIQRYVE--GFLPLPTGIELLVHDGSDAGHNAVEQVYNNKFTGSPBELARKYA 234
      283 AGLVNVVLPIDNGTGSWSLKSFPVP-----GPAP----PLQFYPOGPRESVOC 327
      255 DEEDVVDVYLEDPLPGGKGKGDSTEEPRPLFSSHKRPGDPSPFIHVSGRPLVQPHGPRELEG 314
      328 SRVASSLMTFSPGLGAFSPRIEDVFOGEBRLVYEISLQEAALVYGNSPAATTKRYVDC 387
      315 NALVYGGMSFARLRNSSLGLOLVNHFGBRLAYEVSQDAVALYGGHPPAGMOTYLDV 374
      388 GFGMGKTYTPLTRGVDCPYLATYVDWHFLLSQAAPTIRDAFCVFQONOGPLRRHNSDL 447
      375 GMDLGSVTHELAPGIDCPEPATFDLFHYYDADDPVHYPRALCLFEMPICVPLRRHNSN 434
      448 YS---HYEGGLAEVULVYRSMSTLLNYVDWTFVHPGSAIEIRFYATGYISSAFLGAT 504
      435 FKGGFNFYAGLKGQVLYLRTSTYVNYDYIMDFIFYPNGVMEAKHATGYVHATFYTPEG 494
      505 GKGNQVSEHTLTGTVTHSAHFVKVDLDVAGLEWVMAEDVFPVMAVPMSBEHQLOQV 564
      495 LRGGTLNHLNHTLGNITFLVHNRVDLDVAGTKKSFQTLQKLENTINPMSPRRHVVQPTL 554
      565 TRKLEMEBOAFVLSGASAPRYLYLASNHNKWHAPRGYRIQMLSPAGEPLPONSSVARG 624
      555 EGOYQYMEQOAFNRFKRKLPKYLLFTSPOENPGHKRSYRLOHSMADQVLPFGNGEQA 614
      625 FSWERYQLAVTORKEEPPSSSVFNONDPWAPTVDSDFI--NNETIAGKDLVAVMTAGFL 683
      615 ITWARYPLAVTKRESELSSSIYHQNDDPHVPVFEQFIHNHNEIENEDLVAVMTVGF 674
      684 HIHADIPITVTVGVNGVGFELRPYNEFFDEDPSEFSADSTIFYGDDAGACEVNPPLACLP 743
      675 HIRHSEDIRPATIGNSVGFLLRPFNFEPEDPSLASRDTIV--WPRNGPNVYQ--RWIP 731
      744 QAAACAPDLPAFSHG 759
      732 EDRDCSMP--PFSYNG 746
      732 EDRDCSMP--PFSYNG 746

RESULT 5
B41836
amine oxidase (flavin-containing) (EC 1.4.3.4) precursor - Klebsiella pneumoniae
C:Species: Klebsiella pneumoniae
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 02-Sep-2000
C:Accession: B41836
R:Sugino, H.; Sasaki, M.; Azakami, H.; Yamashita, M.; Murooka, Y.
J. Bacteriol. 174, 2485-2492, 1992
A:Title: A monomine-regulated Klebsiella aerogenes operon containing the monomine oxid
A:Reference number: A41836; MID:92210491; PMID:1556068
A:Accession: B41836
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-755 <SUG>
A:Cross-references: GB:D10208; NID:9216721; PIDN:BA01060.1; PID:d1001529; PID:g216723
A:Superfamily: amine oxidase (copper-containing)
C:Keywords: oxidoreductase

Query Match      10.1%; Score 413; DB 2; Length 755;
Best Local Similarity 27.5%; Pred. No. 2,1e-24;
Matches 150; Conservative 75; Mismatches 213; Indels 108; Gaps 21;

      246 ELLVNHKALDPAWITOKV---FYGGRYVSLAQLEQFAGLVNVVLPIDNGTGSWS- 301
      237 EVLKKHGTIDPGKVVVTTPLTVGFEDGK--DGLQO-----DARLLKVVSYLDTGDGNVMAH 289

```

```

      302 -----LKS-----FVPPGPAP-----PLQF--YPOGPREFSVO 326
      290 PLENVAVVDLEAKKIKIEEGVPIVPEPRPYDGRDNAPVAPLEETPEBGKNTTIT 349
      327 GSRVASSLMTFSPGLGAFSPRIEDVFOG---ERLYEISLQEAALVYGNSPAATTT 382
      350 GDTIHQNMNDFHLRLNSRVGPISTVTVNDNGTKRQVMEEGSLGMIIVYGPBDVGWYK 409
      383 RYVDG--FGMGKTYTPLTRGVDCPYLATYVDWHFLLSQAAPTIRDAFCVFQONOGPLR 441
      410 AYLDSDGYGMGTLPSTIVGKDAFSAVNLIDETADYTKPPTIPGAVAIFE----- 461
      442 RHNSDLYSHYFG-----LAETVLVYRSMSTLLNYVDWTFVHPGSAIEIRYATGYIS 496
      462 RYAGPEYKLEMGKPNVSTERRELVYRWISTVGNVYIIFDMVPHDGTIGIDAGATGTA 521
      497 SAEFLGAT-----GKYNQVSEHTLTGTVTHSAHFVKVDLDVAGLEWVMAEDVFEV 547
      522 VKGVLAKTMHDPESAKEDTFRGTLIDHNIVGTTHQHLYNFRLLDLDVGENNTLVAMPPEVK 581
      548 PNAVPSPEHQLORLQVTRKLEMEBOAFVLSGASAPRYLYLASNHS--NKMGHPGGRYI 605
      582 PNT---AGPRFTSMQVNOYTTIDSEQKAOKEDPGTIR--TLNSTSKENRMGNPVSY-- 633
      606 QMLSFAG--EPLPQNSMA-----RGFWERYQLAVTORKEE-----PSSSVFNG 650
      634 QIIPYAGTHPATGAKFAPDEVIYRLSPMKQLWYTRKHPERPECKYPRRSADHG 693
      651 NDEWAPTVDSDFINNETIAGKDLVAVMTAGFLHIHADIPNTVTVGVNGVGFELRPYNE 710
      694 LQGYAK-----DDESLTNHDDVVMITGTTHVAREBMPIMPT--EWALLALKPMWF 743
      711 FDEDP 716
      744 FDEPT 749

RESULT 6
E64889
amine oxidase (copper-containing) (EC 1.4.3.6) tyra precursor - Escherichia coli (str
N:Alternate names: monomine oxidase; tyraminase; tyramine oxidase
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: E64889; I40923; S65442
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MID:97426617; PMID:9278503
A:Accession: E64889
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-757 <BLAT>
A:Cross-references: GB:AE000235; GB:U00096; NID:91787643; PIDN:AACT4468.1; PID:g17876
A:Experimental source: strain K-12, substrain MG1655
R:Azakami, H.; Yamashita, M.; Roh, J.
J. Ferment. Bioproc. 77, 315-319, 1994
A:Title: Nucleotide sequence of the gene for monomine oxidase (maoA) from Escherich
A:Reference number: I40923
A:Accession: I40923
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-247,'E',249-257,'V',260-276,'I',277-287,289,'I',291-455,'P',457-658,'D
A:Cross-references: GB:D23670; NID:9474301; PIDN:BA04900.1; PID:g809499
A:Experimental source: strain K-12, substrain W3110
R:Steinbach, V.; Benen, J.A.E.; Bader, R.; Postma, P.W.; de Vries, S.; Duine, J.A.
Eur. J. Biochem. 237, 584-591, 1996
A:Title: Cloning of the maoA gene that encodes aromatic amine oxidase of Escherichia
A:Reference number: S65442; MID:96235221; PMID:8647101
A:Accession: S65442
A:Molecule type: protein
A:Residues: 41-48 <STE>

```

A:Experimental source: strain W3350

C:Genetics:

A:Gene: cypA: ma0A

C:Complex: homodimer

C:Function:

A:Description: catalyzes oxidation of phenylethylamine and water to phenylacetaldehyde, A:Pathway: amino acid metabolism; phenylethylamine catabolism
A:Note: 2,4,5-trihydroxyphenylalanine quinone (topaquinone) cofactor; copper cofactor; h
C:Superfamily: amine oxidase (copper-containing)
C:Keywords: copper; oxidoreductase; quinoprotein; topaquinone
F:1-33/Domain: signal sequence #status predicted <STC>
F:34-757/Product: amine oxidase (copper-containing) #status predicted <MAT>

Query Match 9.7%: Score 397.5; DB 2; Length 757;

Best Local Similarity 27.5%: Pred. No. 3,6e-23;

Matches 142; Conservative 74; Mismatches 211; Indels 89; Gaps 20;

```
OY 265 FYGGRYDLSAQLEAFEGALVNVVLIIPDNGTGSMS-----LKS 304
Db 259 YEDGK--DGLKQ-----DRLRLKVLISYLDVGDGNTVANHPIENVAVVLDLQKKIVKIEEG 311
OY 305 PVPKPG-----AP---PLOGY--PGGPRFSVGSRYVASLTFSEFGLGAFSGPR 348
Db 312 PVVPVPMATARPFGDRVARVAKPMQIIEEGKNYITITGDMIMRMMDFLSNKSVGPM 371
OY 349 IEDVRQG-----ERLYEISLDEALAIYGSNSAAMTKRVGCG-FGKGYTTPRLRGVD 403
Db 372 ISTVYVNDNGTKRKVMYEGSLGMIYVGDPIGMWYFKAYLDGSDYGMGTLPRIARXKD 431
OY 404 CPYLATYVDMHFLLESQAKTIRDAFCVEQONGRLPLRHHSDLYSHYFGG-----LAET 458
Db 432 ARSNAYLLMETIADYTGVMETIRATAVE-----KAGREYKHQEGQGVNSTER 483
OY 459 VLVYRSMSTLLNDYVDVTFVHPSGAIEIRFYATGY-----ISSARLFGATGK---YGN 509
Db 484 ELVYRSMSTLLNDYVDVTFVHPSGAIEIRFYATGY-----ISSARLFGATGK---YGN 509
OY 510 QVSEHNLGTVHTSHAFKVDLDVAGLENNVWMAEDMVFVPAVWSPRHQORQVTRKLL 569
Db 544 LIDHNLGTVHTSHAFKVDLDVAGLENNVWMAEDMVFVPAVWSPRHQORQVTRKLL 569
OY 570 EMEEOAFLVGSATPRYLTLASNSNMKGHPRGYRIOMLSFAG--BPLPONSMA----- 622
Db 601 GNDQDAQKFRDPTIR-LLSNPKENRMGNPVSY--QIIYAGCTHNVAKGAPARDEMI 657
OY 623 -RGFSNERYQLAVTORKEEPPSSSVF--NONDPWAPTVDSDFINNETIAGKDLVAVNTA 660
Db 658 YHRLSPNDKQLVWTRVYHGRERPEKGYPNRSTHDTGLGQYSK--DNESLNDTDAVVMYMT 715
OY 681 GFLHHPHADIPNTVTVGVNGVGFLLRPNYFDEDDPS 716
Db 716 GTTHVARAEWPIPT--EWVHTLLKPMWNEDEPT 749
```

RESULT 7

AS6102

amine oxidase (copper-containing) (EC 1.4.3.6) - Arthrobacter globiformis

C:Species: Arthrobacter globiformis

C:Date: 13-Apr-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000

C:Accession: AS6102

R:Choi, Y.H.; Matsuzaki, R.; Fukui, T.; Shimizu, E.; Yorifuji, T.; Sato, H.; Ozaki, Y.;

J. Biol. Chem. 270, 4712-4720, 1995

A:Title: Copper/topa quinone-containing histamine oxidase from Arthrobacter globiformis.

A:Reference number: AS6102; MUID:95181469; PMID:7876243

A:Accession: AS6102

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-684 <CHO>

A:Cross-references: GB:D38508; NID:9994746; PIDN:BA07517.1; PID:9994747

A:Note: authors translated the codon TGC for residue 239 as Phe

C:Superfamily: amine oxidase (copper-containing)

C:Keywords: oxidoreductase

Query Match 9.2%: Score 377; DB 2; Length 684;

Best Local Similarity 24.1%: Pred. No. 1.3e-21;

Matches 182; Conservative 107; Mismatches 343; Indels 124; Gaps 28;

```
OY 39 PHCPVSFSAQPMPT--HPQSQLEADLSREELAVKRFVLRQGLVDAQAQAPSDNVC 96
Db 7 PSLPLVQDPPVPAITLVHAAQHPLEQLSABEIEARILAE--AGLVG----- 52
OY 97 FSVLELDPPKAALAHNDGSP-----PAREALAYFEGROPQ-----PNVSE 140
Db 53 -----ESTFATLGLIEPKTRQGVNQAARLVRLMDAQSRLDVRSLATG 103
OY 141 LVVGLRPHPSYKADVTVERHGGPLRYHRRPVLEFOEYLDIDOMTFNRELPAQSLNHCCF 200
Db 104 LVVD-----RRELNPADG-----QLPVLLEEGGIEDIL--SEDPQMAAL----- 143
OY 201 YKHGRNLTMTAP-----RGLOGDRATMFGLYUNYSGAFELHHVGLLELVNHHKA 253
Db 144 -TARGLTPAQQVAPLSAGVFEYGENEGKRL-----LRGLGFRODP-----ADHPV 189
OY 254 LDPARWTQKVFYQGRYDLSAQLEAFEGALVNVVLIIPD--NGTGSWSLKSVPVPGAP 312
Db 190 AHPIDGLVAFVDVENRRVNL-----IDDGVPV--PVRNGYTPDPAIRGELRTDLR 240
OY 313 PLOGYPOGPRFSVGSRYVASLTFSEFGLGAFSGPRFEDVR--FOGER--LVYEISLQEA 368
Db 241 EIMQEGPSFTLEGHLSAAGMDLRVGFARGLVLIHQHSHKGRRRVIRHASTSEM 300
OY 369 LAIYGGNSPAAMTTRYVCG--FGKGYTTPRLRGVDCPYATYVDMHFLLESQAKTIRD 427
Db 301 VVYEGDPSRYRSMQNFDSSEILVGRDANSLRLGCGDGLITMSPVVADFGPRITEN 360
OY 428 AFGVEEQNGRLPLRHHSDLYSHYFGSLAETLVVYRSMSTLLNDYVDVTFVHPSGAIEI 487
Db 361 GICIHEDDAGILWK--HTDEMAGSDEYRRRRRLVVSFTTVGVNDYGFYWTLYLDGTIEF 418
OY 488 RFATGYISSAFLEGAAGKGNQVSEHTLCTVHTSHAFKVDLDVAGLENNVWMAEDMVFV 547
Db 419 EAKATGIVFTAAALRDKDYAAASELAPGLARYQNHLSAALDMMIDIDARVEELDVRL 478
OY 548 PMAVPMSPHQLORLQVTRKLLMEEOAFLVGSATPRYLTLASNSH--NKGHPRGYRIQ 606
Db 479 PKG--PGKPHG--NAFTQKRLILARESEAVVADACAKKRWNIHNSPDSLNLGHVGYTLY 535
OY 607 MLSFAGEP---LPONSSMARGFSWERYQLAVTORKEEPPSSSVFNONDPWAPTVDSDF 663
Db 536 P---EGNPLAMADDSIASRAAFARHNLWTRHAEELVLAAGDFVNHGGAVALP--AVY 591
OY 664 INNETIAGKDLVAVNTAGFLHHPHADIPNTVTVGVNGVGFLLRPNYFDEDDPSYSTSDI 723
Db 592 AQRDIDGODLVVWHSFGLTHFRPEDWP--IMPVDVTGLTKHKGFEEDNPILVPSA 649
OY 724 YFRGDDAGACEVNPILCLPQAAACAPDLPAPFSHG 759
Db 650 -----AGHCGTG-----SERENHAPGCTAVGHSG 673
```

RESULT 8

G90330

amine oxidase (copper-containing) (cypA) [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001

C:Accession: G90330

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.V.;

Jong, I.; Jeffries, A.C.; Kosem, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redd

arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.

submitted to Genbank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: G90330

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-660 <KUR>
A:Cross-references: GB:AE006641; NID:g13814942; PIDN:AAK11902.1; GSPDB:GN00155
C:Gene: tyna
C:Superfamily: amine oxidase (copper-containing)

Query Match 9.0%; Score 368.5; DB 2; Length 660;
Best Local Similarity 24.4%; Pred. No. 5,8e-21;
Matches 175; Conservative 103; Mismatches 303; Indels 137; Gaps 28;

```
OY 63 LSREELAVNRFLTQRLGPGLVDAQAARPSDNCVSEVLQPPKAAALAHLDKRSPPPAR 122
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 17 LNEEIKRSSEVLRKQLN---LDAKVK-----FFSEVLEKPKKOEYLEMNNKNIKIER 68

OY 123 EALAVFEGQOPNPVSELYVGPLPHPSYMDVVERHNGELPHRRRVLYFOEVLIDDM 182
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 69 ESL-VKYNFQDR-KVYEALTI-----SLDNVVKKEIRSDAAY--PPTLDEGECECKA 118

OY 183 IFNRELPOASGLHHCFCFYKRGKRNLYMTTAPRGLOSGDRATWGLYVYNSGAGFFLH 242
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 119 VRNDRKRVQEA-----LTKRGILINDLNT---LWVDCMARG-----H 151

OY 243 VGEELVHNKALDPAKRTIQKVFYQGRYDSLAOLEAOFEGALVNVLLIPDNGTGSWSL 302
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 152 VDEELRGRRAVIG-YMM-VKKDIEDNGYGRPVHGLMPVVDLKKMEVIRIDHGT----- 203

OY 303 KSPVPPGPA-----PPLQF-YPOGRPSVQGSRAVSSLTWTSFGLGAFSG 346
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 204 -SPRLDADANTYPEKLRKIFGDDLPLEIRQLPSSSIKINMEISWYMRRLRIGTTPREG 262

OY 347 PRIDVFRQGER-----LVYEISLOEALATIGNSPAAMTTRVYDGG-FGKGKATTPLTR 400
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 263 LVITDVKRIDENNRERMLYASVVDLAVPYGDPSPFNKKKMLDAGCYGLCNFLVPLSH 322

OY 401 G-----VDC-PYLATVYDMHFLLESQAPKTIRDAFCVFEQNOGLPLRRHSHDLYSHYG 453
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 323 GNYDLYNDCGEVYIHDYDTRVSSNGTPRIKIKACVHEDEFGVLMR--HTDLRS----- 376

OY 454 GLATF-----VLVYRSMSTLLNYDVMPTVPHPSAIEIRFATGYSIAFLF--GATGKY 507
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 377 GKSEVRNRRLRYLVSFWMTATLAYDGFYFVODGSIEFLVLTGITINDSISEKDPPTKY 436

OY 508 GNOVSEHTLGTVHTSAHFYVDLDVAGLENNVMAEDMVFVMAVPSPEHQLQRLQVTRK 567
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 437 GTRVTPREYVAPIHIFRIRININVDCLRNRY-----EVLNRGEPITREK 481

OY 568 -----LLEMEQAAFLVGSATPRYLYLAS-NHSHKMGHPRGYRIOMLSEAGEP 614
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 482 NPVGNAFEAEENLEENADARRHVPQGRYWKIVNIQKKNYLGLPVAYRLVPGHNVLPR 541

OY 615 LPOSSMARGFSEWERYOLAVYQKKEEPPSSSVFNQNDPMAPTVDFSD-----FINNETI 669
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 542 LPDDSYARRKCAITNYHLMTVPYNEEERYASGDY-----PYLRANDGLPKYIILKKRSI 594

OY 670 AGKDLVAVMTAGFLIHPAEDIPNTVTYGVNGVGFELRPYNFEDEDPSTYSADSIYFG 727
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 595 VDEDLVYVITGIVHVRVIEDMP--VMPVEMAGRLLIPDGFDDKNPITYLERQLRLING 650
```

RESULT 9

A48646

amine oxidase (copper-containing) (EC 1.4.3.6) - *Arthrobacter* sp. (strain P1)

C:Species: *Arthrobacter* sp.

C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 29-Sep-1999

C:Accession: A48646

R:Zhang, X.; Fuller, J.H.; McIntire, W.S.

J. Bacteriol. 175, 5617-5627, 1993

A:Title: Cloning, sequencing, expression, and regulation of the structural gene for the

hydroxylase.

A:Reference number: A48646; MUID:93374858; PMID:8366046

A:Accession: A48646

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-648 <ZHA>
A:Cross-references: GB:J12990; NID:g289152; PIDN:AAA22074.1; PID:g289153
C:Superfamily: amine oxidase (copper-containing)
C:Key words: oxidoreductase

Query Match 8.9%; Score 363; DB 2; Length 648;
Best Local Similarity 24.0%; Pred. No. 1,6e-20;
Matches 177; Conservative 96; Mismatches 301; Indels 164; Gaps 31;

```
OY 55 GOSQLFADLSREELAVNRFLTQRLGPGLVDAQAARPSDNCVSEVLQPPKAAALAHLD 114
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 12 GVSHPLDPLRSVEIARVAALIKE--CPAAESFR-----FISVELREPSKDL----- 57

OY 115 RGSPPRAREALAIYFFGQOPNPVSELYVGPLPHPSYMDVVERHNGELPHRRRVLYFO 174
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 58 RAGVAVAREADAVLV--DRAQRSEAVVD-----LEAGTVSKILLAEINIDPPEFLD 108

OY 175 EYLDIDQMIFNRELPOASGLHHCFCFYKRGKRNLYMTTAPRGLOSGDRATWGLYVY 231
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 109 EFACEDAC--RKDP-----VIALAKRGLINDLVCFEPMWSGYF 148

OY 232 NISGAGFFLHVGLLELVHNKALD-PAKRTIQK--VFYQGRYDSLAOLEAOFEGALVNV 288
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 149 GEDNEGRL--MRALVFYRDEADSPYAHPIENFIYEF-----DLNAG--KV 191

OY 289 VLIDNGTGSWSLSKSPVPPGPAFLQFY-----POGRPSVQGSRAVSS 333
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 192 VRLIEDQA-----IPVSARGNTLPKRYVGEARTDLKPLNITOPEGASFTVTGNHVTWA 244

OY 334 LMTFSFGLGAFSGPRIFVFRQGE-----RLVYEISLOEALATIGNSPAA-MTTRVYDGG 388
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 245 DMSFRVGTTPREGVLVHLNHLKFKDGVDRPYINRASLSMVPYGDATVAVQAKKNAFDSGE 304

OY 389 FGKGYTTPPLTRGVDCPYLATVYDMHFLLESQAAPKTIRDAFCVFEQNOGLPLRRHSHDLY 448
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 305 YNIGMANSLTLCDCGLCEIKYFQGHVSVDSHGNPTIENALCMHEEDDSI-----LM 356

OY 449 SH--YFGSLAET-----VLVYRSMSTLLNYDVMPTVPHPSAIEIRFATGYSIAFLF 501
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 357 KHFDRREGTAETRNSRKLVISFIATVANYETAFVWHLFLDSDSIEFLVATGISTLAGOLP 416

OY 502 GATGKYGNQVSEHTL-GTVHTSAHFYVDLDVAGLENNVMAEDMVFVMAVPSPEHQL- 559
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 417 GEKNPYGGSLNDDGLYARVHQHMFVNRDFFELDGVKKNVYEVDEMEX-----PENHPT 468

OY 560 -QRLQVTRKLLLEMEQAAFLVGSATPRYLYLASNHS-NKMGHPRGYRIOMLSEAGEP 617
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 469 GTAFVAVDRLLLETQKAIKRTNKAHREMKIANHESKNLVNEPVAYRL-----IPT 519

OY 618 NSSMARGFSEWERYOLAVYQKKEEPPSSSVFNQNDPMAPTVD-----FSDFINNET- 668
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 520 NG-----IQLAA--RDDAYVSKRAQFARNNLMTATIDRTFRPAAGETPNQATGAD 567

OY 669 -----IAGKDLVAVMTAGFLIHPAEDIPNTVTYGVNGVGFELRPYNFEDEDP 716
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 568 DGLHMTQKDRNIYDVTDLVWVYTFEGMHVHVRLEDMP--VMPRONIGFLLEPHGFENQPT 625

OY 717 FYSADSIYFRGDDQAGAC 734
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 626 INLPSTSTQTQGEADTC 643
```

RESULT 10

T48139

copper amine oxidase-like protein - *Arabidopsis thaliana*

N:Alternate names: protein T4C9.130

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Sep-2000

C:Accession: T48139

R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dikse, W.; Stiekema, W.; Bancroft, I.;

submitted to the Protein Sequence Database, June 1999

A:Reference number: 224485

A:Accession: T48139

A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-756 <BEV>
 A:Cross-references: EMBL:AL080318
 A:Experimental source: cultivar Columbia; BAC clone T4C9
 A:Genetics:
 A:Map position: 4
 A:introns: 460/3; 498/3; 686/2
 A:Note: T4C9.130
 C:Superfamily: amine oxidase (copper-containing)

Query Match 8.7%; Score 357.5; DB 2; Length 756;
 Best local similarity 22.2%; Pred. No. 5.3e-20;
 Matches 163; Conservative 100; Mismatches 280; Indels 191; Gaps 29;

```

QY 94 NCVEVELQPPKAAALAHLDGSPPPAREALAI-----VFGRQPPQPVSE 140
DB 126 HALHTVLEPERK-NLYRHKKNPLPRKASYIAKVGADTHVLTVDISTGRVDSERSPV 184
QY 141 LVVG-PLPHPSYMRDVTVERHGSLPYHRRPVLFOEYLDIDOMIENRELPOASGLLHCC 199
DB 185 RVSGYPMWTIEEMNDITV-----VPSNAD-----FNRTIIS----- 216
QY 200 FYHNRGRNLVTMTAPRGLOSGRATMFC-----LYNISGAGFLIHHV-GLE 246
DB 217 -----KGVNLDVICFP--ISCG---WFGKKEENARYIKSQCFMTOCTPFNHRPIEGLT 266
QY 247 LVNHRKALDPARMTIQKVFQGRYDLSAQLAEAGLVNVVLPDNGTGSMSLSKSPV 306
DB 267 ILID-----LDTKQ-----VEIIDTGA-----IP1 288
QY 307 PPG-----PAPLPQF-TPQGRFVSQGRVAS-SLMTSEGLGARSGP 347
DB 289 PGSTNTDYRFQKLTATDKTRPLNPISIEOPRGSEFVEDHNLVKNAMERHLKPDPRAGV 348
QY 348 RIDVR-----FOGERLVEISLOEALAYGGNSPAAMTTRYVDG-FGMGKYTPPLTR 400
DB 349 VIRRVAVHDPDTHETRDVAKGFSVSELPYMPDSAMYKTYMDAGEYFGLOAMPVLP 408
QY 401 GVDCPYLATVVDHFFLESQAPKTIRDAFCVFQON-----QGLPLRRHNSDL 447
DB 409 LNDGPRNAAYMGVFAAAGTPEVRNMCIFESYAGDGMHSESPITGIPREVR--- 465
QY 448 YSHFEGGLAETVLYVRSMSLTLNDYWDVTFPHSGAIEIRFYATGITSAPLEGATGKY 507
DB 466 -----PKVTLVVRMAASVGNVDYIDYEFQTDGLIKAKVGLSGIL---MYKGTYYQN 514
QY 508 GNOV-----SEHTIGTVHTHSAHFKYVDLVAGLENNVMAEDMVFVMAV 551
DB 515 KNOVEDKDKGNEELHGLTLESENVIGVHDIHYTYFYLDDVDGPDN-----SFYKVALKR 569
QY 552 PMSPEHOLQ--LOVTRKLLMEEOQAFLVGSATP-RYLYLASNHSNKGHPGRYIOM 607
DB 570 OETEPGSPRSXYLKAARNIAKTEKQDQIKLSTLDPSEFVINSKGTTRGNGPFGYVVP 629
QY 608 LSTRAGEPLPONSMAKFSERYQLAVTQKKEEPPSSSVFNONDPAPIYVD-FSDINN 666
DB 630 RTTAASTLDDHDPQKGAFTNNQIWTVPYKSEOMAGGFTYQSHDDELAVWASD--RD 687
QY 667 ETIAGKDLVAMVTAGFLIHAEIDPNTVGNQVGFELPRPYNFEDDPSFYSDSIYFR 726
DB 688 RDLKNDIVVWYTLGFHIFHCQEDFPIMPTVSS--SFDLAPVNFEEKNPILSAAPN--- 742
QY 727 GDQDAGACEVNPPLA 740
DB 743 -EHDLPVCGVGSVS 755

```

RESULT 11
 G71412
 Probable amine oxidase - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 A:Variety: Columbia

C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 20-Jun-2000
 C:Accession: G71412
 R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.;
 P.; Medler, H.; Medler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.;
 Avaragh, T.; Hempel, S.; Kotter, P.; Entlian, K.D.; Rieger, M.; Schaeffer, M.; Funk,
 Nature 391, 485-488, 1998
 A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Putgdon
 et al.; A.; Moore, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.
 C.; Chaltz, N.
 A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis
 A:Accession: G71412
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-650 <BEV>
 A:Cross-references: GB:297337; NID:G2244829; PID:G2244851
 C:Genetics:
 A:Map position: 4COP9-4G3845
 C:Superfamily: amine oxidase (copper-containing)

Query Match 8.6%; Score 353.5; DB 2; Length 650;
 Best local similarity 23.9%; Pred. No. 8.9e-20;
 Matches 161; Conservative 96; Mismatches 278; Indels 139; Gaps 29;

```

QY 99 VELQLPKAAALAHL-----DRGSPPARALAYFFGRQPPQPVSELVGPPLPHSYMD 154
DB 56 LDLEPPKSHVLOMLSPNPKRPPRRSFVVVRAGQ---TYELI-----D 101
QY 155 VTEVER-----HGP-----LPYHRRPVLFOEYLDIDOMIENREL-POASGLLHCC 199
DB 102 LITSKLAASRIYTHGHPSTFIELFKASLPLTPPFKSLIDRLSINISEVCIPRTVG 161
QY 200 FYHNRGRNLVTMTAPRGLOSGRATMFCGLYVINSAGAFGLHHV-GLELLVNRKALDPA 258
DB 162 WYGE-----TTTRELKA-----SCFYHDSVNVTPRIEGITVTD-ID--- 200
QY 259 WTIQKVFYQGRYDLSAQLAEAGLVNVVLPDNGTGSMSLSKSPVPPQAPLOFY- 317
DB 201 -SMQVIRKYSDFRKP-----IPDK-EGNDFRKH-----RPPFFC 234
QY 318 -PGGRFVSQGRVASSLMTFSFGLAGSPRIFDV-----FOGERLVEISLOEA 368
DB 235 NVSDTGKILGNKVKAMKMFHGFTRAGVTTSTASVLDPRKRRF-RVMTRGHVSER 292
QY 369 LAIYGNSPAAMTTRYVD-GFGMGKYTPPLTRGVDCPYLATYVDWHFFLESQAPKTIR 427
DB 293 FVPMPTPEMYRTFMDIGFGRSAVNLQPLDPCPNAAFLDGHWAGPDGTAQMTN 352
QY 428 ACFVFEQNOGLPLRRHNSDL---SHYFGLAEVLVYRSMSTLLNDYWDVTFPHSGA 484
DB 353 VMCVFEKN-GYGASFRTETINPQOVITSEAEISLVAVATLGNVDYVDWEFKNGA 411
QY 485 IEI-----RYATGYISSAFEGATGGYGNQVSEHTIGVTHSAHFKVDLDVAGLE 536
DB 412 IRYGVDLTGVEYKATSYTNDQ--TENVYGLVANTANTAVNHDHLYLTDLDVDCNG 469
QY 537 NMYAEDMVFVMA-VMSPEHOLQLOVTRKLLMEEOQAFLVGSATPRLYLASHNS 595
DB 470 NSLVKAKLKTVRVTEVKNKTSRSKSYVTYKETADEADRVRGSDPVELLYNPKKT 529
QY 596 KMGHPRGRYIOMLSFAGEPLP---QNSMARGFSEMYQLAVYQKKEEPPSSSVFNOND 652
DB 530 KIGNTGYRL-----IPEHLQATSLTDDYPELRAGY---KYPVWYATYDSE 576
QY 653 PMAPTVDS-----FINNETIAGDLVAMVTAGFLIHAEIDPNTVYNGV 701
DB 577 RMAGGF-YDSRSKGDGDLAVWSSNNRLEKNKDIYMTYNGFHHIYQEDDEPVPMTLNG-- 633
QY 702 GFELRPYNFDEDP 715
DB 634 GFTLRPSNFDNDP 647

```


Db 545 EYGYRL-ITAIPIAHPLLEDDPYQ-----IRG-AFTNYNWT-----AYNRTEK 587

Qy 654 WAP--TVDES---DFI-----NNETIAGDLVAMVATAGFLHIPHADIPTNTVGVNGVGF 703

Db 588 MAGGLYVDHSDRDDLAVMTKONREIVNKDIYMHVYGIHVPAGDDPIMPLLL--STSF 645

Qy 704 FLRPYNEFEDP 715

Db 646 ELRPNTFEFRNP 657

RESULT 14

T47403

amine oxidase-like protein - Arabidopsis thaliana

N:Alternate names: protein F23N14.50

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 21-Jul-2000

C:Accession: T47403

R:Blocher, H.; Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.; Quetler, F.; Salanoubat

submitted to the Protein Sequence Database, March 2000

A:Reference number: 224465

A:Accession: T47403

A:Status: preliminary

A:Molecule type: DNA

A:Cross-references: EMBL:AL138638

A:Experimental source: cultivar Columbia; BAC clone F23N14

C:Genetics:

A:Map position: 3

A:Intons: 145/3; 303/1; 402/3; 440/3; 591/3; 624/2

A>Note: F23N14.50

C:Superfamily: amine oxidase (copper-containing)

Query Match 8.0%; Score 327; DB 2; Length 687;
 Best Local Similarity 21.6%; Pred. No. 1.2e-17;
 Matches 171; Conservative 113; Mismatches 314; Indels 194; Gaps 31;

Qy 7 LVLLIAVITIPALVGVLVGRGDGGEPSQLPHCPVSFSAQPMWHPGSOQLFADLSRE 66

Db 9 LVLLISLFLFTTL-----ASSKTRFKYSLEKPHHP---LDPLTTP 49

Qy 67 ELTAIVARFLTORLPGVLVDAQAQPSDNCVSEVLEQLPRKAALAHLDGSPPARALA 126

Db 50 EIRKVTITLSGH-DPFGSGS-----TIHMALEDPRKORIRKKGRLPRRAEIL 102

Qy 127 IYFGKQOPQV---SELVVGRLPHPSY-----MDVTVRHHGGLPYRRARVLEQYL 177

Db 103 AMNGESHVLTVDLKSGRVYSDLVNPTFGYPIITMKDIIV---SQVPY-----KSV 151

Qy 178 DIOQMFPNRELPOASGLIHHCCYKRRGLVMTTAAPKLGSGDRATWGLYNNISGAG 237

Db 152 EFRRSIEARQIP-FSGIICITPPAGWYGPD-----EGGRARYIKIQCFSKODTVN 199

Qy 238 FELAHY-GLELLVNHKAALDPAWTIQKVFQGRYYSLAQLKQAEKGLNVVLPINGT 296

Db 200 FYKRPLEGILYVDMKLE-----TIKIYDNG- 226

Qy 297 GGSMSLSKSPVPPAPAPLQF-----YQGPREFSVQ-GSRVASSLMTF 337

Db 227 -----PVYPKSTGEYRGLFENETVYMDRVNPMSEQDPSFOVEDYILVKMANMKF 280

Qy 338 SEGLAFSGPRITVDNRFQGERLYEISLOEALAI-YGNSPAA-----TTRYV 385

Db 281 HIKPDORAGMI-----SOATVRSKTGEARVMTKGFASELFPVMDPEGEGYSKAYM 334

Qy 386 DGG-FGMAKTYTLRGVNDPRLATYVDMHLEESQAPKTRDAFCVEFQNOGLPLRRH 444

Db 335 DADEFGLGSSMPLVPLNDPRLNAYITDGFASPEGIPILQPNNICLEERYADQTSWRHS 394

Qy 445 SDLYSHYFGGL-----AETLVVRSMTSLNTLDYVDVTFHPSGAIEIRFYATG----- 493

Db 395 EIL-----LPGVDIRESAKATLVARMAAGSGVNDYIFIDMEFQNDGIRLVYVAASGMLMK 450

Qy 494 ---YISSAFLEGATGTYGNOVSEHTLGYTHSAHKVLDVAGLENNWMAEDMVFVMA 550

Db 451 GYATENEVEDLGERKEDDSGLISENNYGVYHDEIFSHLMDIDGSAN---NSFYVHLE 506

Qy 551 VPMSPHOLQ---LOVTRKLEEMEQAAFLVGSATPRYLTLAS-HSKNKGHPRGYRIQ 606

Db 507 KORLPGESRRKSYLKKYKTVAKTEKDAQIKNSLYPRFEHLVNPRLSLRGNPAGYKLV 566

Qy 607 MLSFAGEPLPONSSMARGEWMERYQLAVTQRKEEPPSSSVFNQNDPMA----- 655

Db 567 PGNAASLIDHDPPQMRGAFTNQIWTNR-----YRSEOMAGLLMYQSGE 615

Qy 656 PYVD-FSDFINNETINGKLVAMVATAGFLHIPHADIPTNTVGVNGVGFRLRYNFEDD 714

Db 616 DTLQVMSD---RRRSINKDIYLVYTLGFFHVPQDEDFPVMPTIAS--SFLKRVNFEEN 671

Qy 715 P-----SFYSAD 721

Db 672 PVLGISPFREKD 683

RESULT 15

C44239

amine oxidase (copper-containing) (EC 1.4.3.6) precursor - garden pea

C:Species: Pisum sativum (garden pea)

C:Date: 31-Dec-1993 #sequence_revision 05-Jan-1996 #text_change 17-Mar-2000

C:Accession: A57327; C44239

R:Tippling, A.J.; McPherson, M.J.

J. Biol. Chem. 270, 16939-16946, 1995

A:Title: Cloning and molecular analysis of the pea seedling copper amine oxidase.

A:Reference number: A57327; MUID:95348126; PMID:7622512

A:Accession: A57327

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-674 <TIP>

A:Cross-references: GB:J39931; NID:9685197; PID:9685198

R:Jones, S.M.; Palcic, M.M.; Scaman, C.H.; Smith, A.J.; Brown, D.E.; Dooley, D.M.;

Biochemistry 31, 12147-12154, 1992

A:Title: Identification of topaquinine and its consensus sequence in copper amine c

A:Reference number: A44239; MUID:93090748; PMID:1457410

A:Accession: C44239

A:Molecule type: protein

A:Residues: 409-417, 'X', 419 <JAN>

A:Experimental source: seedling

C:Superfamily: amine oxidase (copper-containing)

C:Keywords: copper; glycoprotein; metalloprotein; oxidoreductase; quinoprotein; top

F:1-25/Domain: signal sequence #status predicted <Sig>

F:26-674/Product: amine oxidase (copper-containing) #status predicted <Mat>

F:156-389/Binding site: carbohydrate (asn) (covalent) #status predicted

F:382-467/Binding site: copper (His) #status predicted

F:412/Modified site: topaquinine (Tyr) #status experimental

Query Match 7.8%; Score 320; DB 2; Length 674;
 Best Local Similarity 25.8%; Pred. No. 4.3e-17;
 Matches 120; Conservative 66; Mismatches 189; Indels 90; Gaps 20;

Qy 301 SLKSPVPPGAPP--LQFYQGRFRFSVQSGRVASSLTFSGFGAFSG-----PRIFDV- 352

Db 233 SKQSP--PFGPKQSLTSHQQGFQFQJNGSHVAMNKKPHIGDVRAVGIYISLASIYDL 291

Qy 353 RFGGERLYVEISLOEALAIYGSNPAAMTRYVDG-FGNGKTYTLPLRGVDCPYLATYV 411

Db 292 KKSRLVLYKGYISELFPVQDPTEEFYKTFPDSGEGFGLSVSLIPNRDPPHQAQFI 351

Qy 412 DMHFLAESQAPKTRIDAFVEFQNOGLPLRRHSDLSYSHFGGL-----AETLVV 462

Db 352 DTVVHSANGTPIILKNAICVFEQ-----YGNIMRRHENGIPNIESIESRETVNLIV 403

Qy 463 RSNSTLNDYVDVTFHPSGATEIRFATGYISSAFLEGATG-----YGNQVSE 513

Db 404 RTIVYGVNDVNDIMWEKAGSGIKPSIALSGILE---IKGTNKKHKEIKEDLHGKLVSA 460

Qy 514 HTLGYVTHSAHKVLDVAGLENNWMAEDMVFVPM-----AVPMSPHOLQRLQVTR 566

```

Db 461 NSIGYHDFYLYLDIDGTGHNSEKTSKTVRIKDOSSKRKSWTETOTAK----- 515
OY 567 KLEMEBOAPFLVGSATRIYLYLASNHSNKNGHPRGYRIOMLSTAGEPL-----PQNSS 620
Db 516 ---TESDAKITIGLAPAPELVYVNPNIKTAVGNEVGYRL-IPAIPAHPILTEDDYPO--- 567
OY 621 MARGFSMERYOLAVTORKEEERSSSVFNONDPMAP--TVDFS---DFI-----NNETIA 670
Db 568 -IRG-AFTNINVMW7-----AYNRTEKMAAGLYVDHSRGDDTLAVWTKONREIV 614
OY 671 GKDLVAVWTAGFLAIPHAEDIPNTVTGNGVGFFLRPNFDEDP 715
Db 615 NKDIYMHVVGIIHVPQOEDEFIMPL--STSEFLRPTNFERNP 657

```

Search completed: May 20, 2003, 11:39:22
 Job time : 33.4394 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 08:02:28 ; Search time 18.01 Seconds
(without alignments)
1757.161 Million cell updates/sec

Title: US-10-081-408-2
Perfect score: 4095
Sequence: 1 MNOKTILVLLITAVITFAL.....QAAACAPDLPAFSGFSHN 763

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4095	100.0	763	1 AOC3_HUMAN	Q16853 homo sapien
2	3473	84.8	765	1 AOC3_MOUSE	O70423 mus musc
3	3346	81.7	762	1 AOC3_BOVIN	O46406 bos taur
4	3334	81.4	762	1 AOC3_BOVIN	O28437 bos taur
5	2612	63.8	756	1 AOC3_HUMAN	O71106 homo sap
6	1418	34.6	321	1 AOC3_RAT	O08590 rat
7	1400.5	34.2	746	1 ABP_RAT	P36633 rat
8	1394	34.0	751	1 ABP_HUMAN	P18801 homo sap
9	413	10.1	755	1 AMO_KLEAE	P49250 klebsiella
10	397.5	9.7	757	1 AMO_ECOLI	P46883 escherich
11	377	9.2	683	1 AMO_ARTGO	O59118 arthroba
12	363	8.9	648	1 AMO2_ARTSI	O07121 arthroba
13	363	8.9	648	1 AMO2_ARTSI	O07123 arthroba
14	340	8.3	638	1 PAOX_ARTGO	P46881 arthroba
15	320	7.8	674	1 AMO_PEA	O43077 pisum sat
16	311.5	7.6	666	1 AMO_LENCU	P49252 lens cul
17	310	7.6	671	1 AMO1_ASPNG	O12556 aspergill
18	295	7.2	692	1 AMO_PICAN	P12807 pichia an
19	111	2.7	1322	1 POR4_XYLFA	O96416 xyliella
20	104.5	2.6	736	1 DVL2_MOUSE	O60838 mus musc
21	103	2.5	736	1 DVL2_HUMAN	O16441 mus sap
22	103	2.5	2224	1 PAS_HUMAN	P12259 homo sap
23	102.5	2.5	591	1 COXN_BRAJA	P98000 bradyhiz
24	101.5	2.5	1520	1 ACED_ECOI	O46837 escherich
25	101	2.5	435	1 KICH_MOUSE	O54804 mus musc
26	100.5	2.5	597	1 DIVJ_MOUSE	O03228 cauloba
27	100.5	2.5	1257	1 CAML_HUMAN	P33004 homo sap
28	98.5	2.4	881	1 NIAL_PHAVU	P38665 phaselus
29	98.5	2.4	1217	1 S3B3_HUMAN	O13393 homo sap
30	98	2.4	2212	1 RRPL_ERBOM	O05318 ebola vir
31	97	2.4	716	1 DVL3_HUMAN	O92997 homo sap
32	97	2.4	716	1 DVL3_MOUSE	O61062 mus musc
33	97	2.4	733	1 HEXA_BIADI	O17127 blaiberu

ALIGNMENTS

RESULT 1	ID	ACOC3_HUMAN	STANDARD:	PRT:	763 AA.
AC	O16853:				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Membrane copper amine oxidase (EC 1.4.3.6) (Vascular adhesion protein-1) (VAP-1) (HRAO).				
DE	1) (VAP-1) (HRAO).				
GN	AOC3 OR VAP1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Placenta;				
RX	MEDLINE=97128319; PubMed=8972912;				
RA	Zhang X., McIntire W.S.;				
RT	"Cloning and sequencing of a copper-containing, topaquinine-				
RT	containing monoamine oxidase from human placenta."				
RL	Gene 179:279-286(1996).				
RN	[2]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	TISSUE=Lung;				
RX	MEDLINE=98317014; PubMed=9653080;				
RA	Smith D.J., Salimi M., Bono P., Hellman J., Leu T., Jalkanen S.;				
RT	"Cloning of vascular adhesion protein 1 reveals a novel				
RT	multifunctional adhesion molecule."				
RL	J. Exp. Med. 188:17-27(1998).				
CC	- FUNCTION: CELL ADHESION PROTEIN THAT PARTICIPATE IN LYMPHOCYTE				
CC	RECIRCULATION BY MEDIATING THE BINDING OF LYMPHOCYTES TO				
CC	PERIPHERAL LYMPH NODE VASCULAR ENDOTHELIAL CELLS IN AN E-SELECTIN-				
CC	INDEPENDENT FASHION. HAS A MONOAMINE OXIDASE ACTIVITY.				
CC	- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +				
CC	H(2)O(2).				
CC	- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.				
CC	- SUBUNIT: HOMODIMER.				
CC	- SUBCELLULAR LOCATION: Type II membrane protein.				
CC	- TISSUE SPECIFICITY: MOST STRONGLY EXPRESSED ON THE HIGH				
CC	ENDOTHELIAL VENULES OF PERIPHERAL LYMPH NODES AND ON HEPATIC				
CC	ENDOTHELIA.				
CC	- PTM: Topaquinine (TPQ) is generated by copper-dependent				
CC	autooxidation of a specific tyrosyl residue (By similarity).				
CC	- PTM: N- AND O-GLYCOSYLATED.				
CC	- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.				
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
DR	EMBL, U39447; AAC50919.1; -				

34 96.5 2.4 1137 1 RIRL_HSV11 P08543 herpes simp
35 96 2.3 421 1 SYS_THETH P34945 thernus the
36 95.5 2.3 553 1 CB62_ARATH O23066 arabidopsi
37 95.5 2.3 981 1 EPA3_BRARE O13146 brachydanto
38 95.5 2.3 985 1 AGLU_ASPEG P56926 aspergillus
39 95.5 2.3 2175 1 POLG_BOVEY P12915 bovine ente
40 95 2.3 2258 1 FAS_PIG O961P1 sus scrofa
41 94.5 2.3 901 1 VEF_GYPU P41723 pseudolatia
42 94.5 2.3 1522 1 PST1_SCHPO O09750 schizosach
43 94.5 2.3 2555 1 PPS3_BACSU P39847 bacillus su
44 94 2.3 492 1 SYTM_YEAST P48527 saccharomyc
45 93.5 2.3 878 1 FIMD_ECOLI P30130 escherichia

```

DR EMBL: AF067406; AAC25170.1; -.
DR Genew: HGNC:550; AOC3.
DR MIM: 603735; -.
DR InterPro: IPR000269; CUNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
DR Pfam: PF02727; Cu_amine_oxidN3; 1.
DR Pfam: PF02728; Cu_amine_oxidN3; 1.
DR PRINTS: PR00766; CUDAOXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
DR Oxioreductase; Copper; TPO; Glycoprotein; Transmembrane;
KM Signal-anchor; Cell adhesion; Polymorphism; Metal-binding.
FT DOMAIN 1 5
FT TRANSMEM 6 26
FT SIGNAL-ANCHOR (TYPE-TI MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 27 763
FT MOD_RES 471 471 TOPAQUINONE (BY SIMILARITY).
FT METAL 520 520 COPPER (POTENTIAL).
FT METAL 522 522 COPPER (POTENTIAL).
FT METAL 684 684 COPPER (POTENTIAL).
FT BINDING 673 673 AMLOXIDE (BY SIMILARITY).
FT CARBOHYD 137 137 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 592 592 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 618 618 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 666 666 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 317 317 Y -> H (IN DBSNP:438287).
FT /FTID-VAR_012064.
SQ SEQUENCE 763 AA; 84621 MW; 58AD55605EC9D228 CRC64;
Query Match 100.0%; Score 4095; DB 1; Length 763;
Best Local Similarity 100.0%; Pred. No. 1.3e-311;
Matches 763; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNOKTLLVLLAVITFALVLCVLLVGRGSGEPPSOLPHCPSPSPSAQPMTHGQSLF 60
DB 1 MNOKTLLVLLAVITFALVLCVLLVGRGSGEPPSOLPHCPSPSPSAQPMTHGQSLF 60
QY 61 ADSREELTAVMRELTORGLVDAQAARPNVCNVSLEOLPRKAAALAHLDGSGPP 120
DB 61 ADSREELTAVMRELTORGLVDAQAARPNVCNVSLEOLPRKAAALAHLDGSGPP 120
QY 121 AREALAVFEGROPNPVSELVGPLPHPSYMRDVYERHGGPLPYRRRNVLFQEYDID 180
DB 121 AREALAVFEGROPNPVSELVGPLPHPSYMRDVYERHGGPLPYRRRNVLFQEYDID 180
QY 181 QMIFNELPQASGLLHHCFFYKRGRLVMTTAPRGLOSGBRATWGLYNNISGAGFL 240
DB 181 QMIFNELPQASGLLHHCFFYKRGRLVMTTAPRGLOSGBRATWGLYNNISGAGFL 240
QY 241 HNYGLELVNHHKALDPAKMTIQVYOGRYDLSLAOLEAFEGALVNVLLIPDNGTGSW 300
DB 241 HNYGLELVNHHKALDPAKMTIQVYOGRYDLSLAOLEAFEGALVNVLLIPDNGTGSW 300
QY 301 SLKSPVPGPAPLOFYPOGPRFSVGSRAVSLMTFSEGLAFSGRITDVRQGRIV 360
DB 301 SLKSPVPGPAPLOFYPOGPRFSVGSRAVSLMTFSEGLAFSGRITDVRQGRIV 360
QY 361 YEISLOEALAIYGSNSPAAMTTRYVDGEGMGKYYTTPRLTRGVDCPYLATYVDHFLLESQ 420
DB 361 YEISLOEALAIYGSNSPAAMTTRYVDGEGMGKYYTTPRLTRGVDCPYLATYVDHFLLESQ 420
QY 421 APTTIDACVFEQONGRLPRRHHSLYSYFEGLAETVLVVRSMSTLLNYDVVMDTVFH 480
DB 421 APTTIDACVFEQONGRLPRRHHSLYSYFEGLAETVLVVRSMSTLLNYDVVMDTVFH 480
QY 481 PSQAIRFATGYISSAFLEGATGKYGNOVSEHTLGTVTHSAHFKVYDIDVAGLEMMV 540
DB 481 PSQAIRFATGYISSAFLEGATGKYGNOVSEHTLGTVTHSAHFKVYDIDVAGLEMMV 540
QY 541 AEDVVEPAVVPSPHQLOLQVTRKLEMEQAAFLVGSATPRYLVLASNSNKGHP 600
DB 541 AEDVVEPAVVPSPHQLOLQVTRKLEMEQAAFLVGSATPRYLVLASNSNKGHP 600

```

```

DB 541 AEDVVEPAVVPSPHQLOLQVTRKLEMEQAAFLVGSATPRYLVLASNSNKGHP 600
QY 601 RGYRIQMLSEFAGEPLLPONSSMARGFSWERYQLAVTORKEEPESSSVFNONDPAPTVDF 660
DB 601 RGYRIQMLSEFAGEPLLPONSSMARGFSWERYQLAVTORKEEPESSSVFNONDPAPTVDF 660
QY 661 SDFINNETIAGKDLVAVYTAGFLIHPAEDIPNTVTVGVNGVGFLLRPYNEFDEDPSTYSA 720
DB 661 SDFINNETIAGKDLVAVYTAGFLIHPAEDIPNTVTVGVNGVGFLLRPYNEFDEDPSTYSA 720
QY 721 DSIFYRGDDQAGACEVNPPLACLQAAAACADLPFAFSGSGSHN 763
DB 721 DSIFYRGDDQAGACEVNPPLACLQAAAACADLPFAFSGSGSHN 763
RESULT 2
AOC3_MOUSE
ID AOC3_MOUSE STANDARD; PRT; 765 AA.
AC 070423;
AC 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Membrane copper amine oxidase (EC 1.4.3.6) (Vascular adhesion protein-
DE 1) (YAP-1)
GN AOC3 OR VAP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/c, and 129/SvJ;
RX MEDLINE=98414290; PubMed=9743358;
RA Bono P., Salmi M., Smith D.J., Leppanen I., Horelli-Kuitunen N.,
RA Palotie A., Jalkanen S.;
RT Isolation, structural characterization, and chromosomal mapping of
RT the mouse vascular adhesion protein-1 gene and promoter.;
J. Immunol. 161:2953-2960(1998).
CC -1- FUNCTION: CELL ADHESION PROTEIN THAT PARTICIPATE IN LYMPHOCYTE
CC RECIRCULATION BY MEDIATING THE BINDING OF LYMPHOCYTES TO
CC PERIPHERAL LYMPH NODE VASCULAR ENDOTHELIAL CELLS IN AN L-SELECTIN-
CC INDEPENDENT FASHION. HAS A MONOAMINE OXIDASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
CC H(2)O(2).
CC -1- CORFACTOR: Binds 1 copper ion and 1 topaquinone per subunit.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
CC -1- PTM: Topaquinone (TPO) is generated by copper-dependent
CC autoxidation of a specific tyrosyl residue (By similarity).
CC -1- PTM: N- AND O-GLYCOSYLATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DB EMBL: AF054831; AAC23747.1; -.
DB EMBL: AF078705; AAC35839.1; -.
DB MGD: MGI:1306797; AOC3.
DR InterPro: IPR000269; CUNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
DR Pfam: PF02727; Cu_amine_oxidN2; 1.
DR Pfam: PF02728; Cu_amine_oxidN3; 1.
DR PRINTS: PR00766; CUDAOXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
KM Oxioreductase; Copper; TPO; Glycoprotein; Transmembrane;
KM Signal-anchor; Cell adhesion; Metal-binding.
FT DOMAIN 1 6
FT CYTOPLASMIC (POTENTIAL).

```

```

FT TRANSMEM 7 27 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL)
FT DOMAIN 28 765 EXTRACELLULAR (POTENTIAL)
FT MOD_RES 471 471 TOPADIOLONE (BY SIMILARITY)
FT METAL 520 520 COPPER (POTENTIAL)
FT METAL 522 522 COPPER (POTENTIAL)
FT METAL 684 684 COPPER (POTENTIAL)
FT BINDING 673 673 AMILORIDE (BY SIMILARITY)
FT CARBOHYD 137 137 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 592 592 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 659 659 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 666 666 N-LINKED (GLCNAC. . .) (POTENTIAL)
SQ SEQUENCE 765 AA; 84533 MW; 7489567D3DB844D CRC64;

Query Match 84.8%; Score 3473; DB 1; Length 765;
Best Local Similarity 83.2%; Pred. No. 4.5e-263;
Matches 634; Conservative 61; Mismatches 67; Indels 0; Gaps 0;

QY 1 MNOKTILVLLIIVITTFALVCLVLRGSGDGEPSQLPHCPSPVSPAPWTHPGSQGLF 60
DB 1 MTKKTLLVLLAVITTFALVCLVLRGSGDGLSLPLCPSPVSPQRTHPSSQPF 60
QY 61 ADLSRELTAVMRFLOKRGGLVDAQAARSPNCVPSVELQIPKRAALAHIDRSGPP 120
DB 61 ADLSPELTAVMRFLOKRGGLVDAQAARSPNCVPSVELQIPKRAALAHIDRSGPP 120
QY 121 ARBALAVFPGROPPOPVSELVGLPLPHPSMDVYERHGRGLPYRHRPRLFOEYLD 180
DB 121 VRELAIIFFGGPKPVPVSLVGLPLPHPSMDVYERHGRGLPYRHRVLDREVODIE 180
QY 181 QMIFNRRLPQASGLLHCCCKYKRGKRLVMTTAPRGLOGSDRATPFGLYINISGAFEL 240
DB 181 EMIFHRLPQASGLLHCCCKYKRGKRLVMTTAPRGLOGSDRATPFGLYINISGAFEL 240
QY 241 HHGGLLVNKKALDPAWMTIOKVEYOGRYDLSLAOLEAEGVNVVLIPIPGNGSGW 300
DB 241 HPIGGLLVNKKALDPAWMTIOKVEYOGRYDLSLAOLEAEGVNVVLIPIPGNGSGW 300
QY 301 SLKSPVPGPAPPLQGYPGCPREFSVSGSRVASSLMTFSGLGAFSGPRIFDVPRGERLV 360
DB 301 SLKSPVPGPAPPLQGYPGCPREFSVSGSRVASSLMTFSGLGAFSGPRIFDVPRGERLV 360
QY 361 YEISLQDALIYGGNSPAAATTRYVDGFGMGKTYTPLTGVCVCPRIATVDMHFLLESQ 420
DB 361 YEISVQDALIYGGNSPAAATTRYVDGFGMGKTYTPLTGVCVCPRIATVDMHFLLESQ 420
QY 421 APKTIKDAFCVFNQGLPLRRHSDLYSHYFGLAETVLVVMSMTLLNVYDMVDTVFH 480
DB 421 APKTIKDAFCVFNQGLPLRRHSDLYSHYFGLAETVLVVMSMTLLNVYDMVDTVFH 480
QY 481 PSQAIEIRYATGYISSAFLEGATGKXGNOVSEHTLGTVTHSHAFKVDLDVAGLENVW 540
DB 481 PSQAIEIRYATGYISSAFLEGATGKXGNOVSEHTLGTVTHSHAFKVDLDVAGLENVW 540
QY 541 AEMNVVPMAPVWSPHQLOLQVTRKLLMEFOAALVLSAIPRYLYLASNSNKKGHR 600
DB 541 AEMNVVPMAPVWSPHQLOLQVTRKLLMEFOAALVLSAIPRYLYLASNSNKKGHR 600
QY 601 RGYRIOMLSPAGEPLPONSMMARGFSEWERYQLAVTORKEEPPSSSVFNONDPAAPYVDF 660
DB 601 RGYRIOMLSPAGEPLPONSMMARGFSEWERYQLAVTORKEEPPSSSVFNONDPAAPYVDF 660
QY 661 SDFINNETTAGKDLVAVNTAGTLHIPHAEDIPMTVTVGVNGVGFGLRPNFDEDEPSFYSA 720
DB 661 SDFINNETTAGKDLVAVNTAGTLHIPHAEDIPMTVTVGVNGVGFGLRPNFDEDEPSFYSA 720
QY 721 DSIFYREGDODAGCEVNPCLCLPQAACARDLPAFSGSGSH 762
DB 721 DSIFYREGDODAGCEVNPCLCLPQAACARDLPAFSGSGSH 762
QY 721 DSIFYREGDODAGCEVNPCLCLPQAACARDLPAFSGSGSH 762
DB 721 DSIFYREGDODAGCEVNPCLCLPQAACARDLPAFSGSGSH 762

```

```

RESULT 3
ACCT_BOVIN
ID ACCT_BOVIN STANDARD: PRT: 762 AA.
AC 046406;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Copper amine oxidase, lung isozyme precursor (EC 1.4.3.6) (Amine
OS oxidase [copper-containing]) (BOLA0).
OC Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_Taxid=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=98151264; PubMed=9492300;
RA Hogdell E.V.S., Huen G., Borre M., Bundgaard J.R., Larsson L.-I.,
RA Vuust J.;
RT "Structure and tissue-specific expression of genes encoding bovine
RT copper amine oxidases."
RL Eur. J. Biochem. 251:320-328(1998).
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +
CC H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit (By
CC similarity).
CC -1- ENZYME REGULATION: INHIBITED BY AMILORIDE IN A COMPETITIVE MANNER
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Extracellular (By similarity).
CC -1- TISSUE SPECIFICITY: LUNG, SPLEEN, HEART, KIDNEY.
CC -1- PTM: Topaquinine (TPQ) is generated by copper-dependent
CC autooxidation of a specific tyrosyl residue (By similarity).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPADIOLONE OXIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC
CC EMBL: Y15774; CA75776.1; -
CC InterPro: IPR000269; CuNH_oxidase.
CC Pfam: PF01179; Cu_amine_oxid_1.
CC Pfam: PF02728; Cu_amine_oxidn3; 1.
CC PRINTS: PR00766; CUAMOXIDASE.
CC PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
CC PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
CC Oxidoreductase; Copper; TPQ; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 762 COPPER AMINE OXIDASE, LUNG ISOZYME.
FT MOD_RES 470 470 TOPADIOLONE (BY SIMILARITY).
FT METAL 519 519 COPPER (POTENTIAL).
FT METAL 521 521 COPPER (POTENTIAL).
FT METAL 521 521 COPPER (POTENTIAL).
FT METAL 683 683 AMILORIDE (BY SIMILARITY).
FT BINDING 672 672 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 617 617 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 665 665 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 762 AA; 84883 MW; BB43D04776744AF2 CRC64;

Query Match 81.9%; Score 3346; DB 1; Length 762;
Best Local Similarity 81.9%; Pred. No. 3.6e-253;
Matches 620; Conservative 46; Mismatches 81; Indels 10; Gaps 2;

QY 14 VITIFALVCLVLR--GGDGE-----PSQLPHCPSPVSPAPWTHPGSQGLFADL 63
DB 3 IFIFLVLVLLVNGREGGVGSEGVGKQCHPSLPKPCSPSDQWTHPDQSQGLFADL 62

```



```

OY 64 SRELTAVMFLTORLGRGLVDAQAARPSDNCVSEVLOLPPKAAALAHLDKSSPPARE 123
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 63 SRELTAVMFLTORLGRGLVDAQAARPSDNCVSEVLOLPPKAAALAHLDKSSPPARE 122
OY 124 ALAIVFGROPORNVSELVVGPLPHPSYMDVYVERHGGLPYRRRPVLFQEYLDIDOMI 183
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 123 ALAIVFGROPORNVSELVVGPLPHPSYMDVYVERHGGLPYRRRPVLFQEYLDIDOMI 182
OY 184 FNNELPOASGLLHHCCEYKRGKRLVMTTAPRGLSGDRATWFGLYYNSGAGFLAHV 243
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 183 FNNELPOAAGVLLHHCCEYKRGKRLVMTTAPRGLSGDRATWFGLYYNSGAGFLAHV 242
OY 244 GLELLVNHKALDPAKMTIOKVFYQGRYDSLAOLEAFGLVNVVLIIPNGTGSMSLK 303
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 243 GLELLVNHKALDPAKMTIOKVFYQGRYDSLAOLEAFGLVNVVLIIPNGTGSMSLK 302
OY 304 SPVPGPAPPLQFYPOGPRFSVQGSRAVSSSLMTFSEGLGAFSGPRIPDFVQERLYEYI 363
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 303 SPVPGPAPPLQFYPOGPRFSVQGSRAVSSSLMTFSEGLGAFSGPRIPDFVQERLYEYI 362
OY 364 SLEALAIYGSNSPAAATTTVYVQGGFMGKXTPTLTGVCPCPYLATYVDHMFLESQAPK 423
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 363 SLEALAIYGSNSPAAATTTVYVQGGFMGKXTPTLTGVCPCPYLATYVDHMFLESQAPK 422
OY 424 TINDAFCEPQONGLPLRRHSDLSHYFGSLAETVLVRSMSSTLNVYDYVMDTVFHPSG 483
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 423 TINDAFCEPQONGLPLRRHSDLSHYFGSLAETVLVRSMSSTLNVYDYVMDTVFHPSG 482
OY 484 AIEIRYATYISAFLEGATGYGNOVSEHTLTGTVTHSAHFVKDLDVAGLEMMVAED 543
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 483 AIEIRYATYISAFLEGATGYGNOVSEHTLTGTVTHSAHFVKDLDVAGLEMMVAED 542
OY 544 MIVVPVAPVPSPEHOLQLOVTRKLEMEQAAFLVGSAPRITYLXLSNNSKNGHGRGY 603
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 543 MIVVPVAPVPSPEHOLQLOVTRKLEMEQAAFLVGSAPRITYLXLSNNSKNGHGRGY 602
OY 604 RIOMLSFAGPPLPONSSMANGFSEWERYQALAVTORKEEPPSSSVFNONDPAATVDPDSF 663
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 603 RIOMLSFAGPPLPONSSMANGFSEWERYQALAVTORKEEPPSSSVFNONDPAATVDPDSF 662
OY 664 INNETTAGKLVAVMTAGFLHIHPAEDIPNTVVGNGVGFELRPYNEFDEDPFSYASDI 723
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 663 INNETTAGKLVAVMTAGFLHIHPAEDIPNTVVGNGVGFELRPYNEFDEDPFSYASDI 722
OY 724 YFRGDODAGCEVNHACLPOAAACAPDLPAFSGGF 760
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 723 YFRGDODAGCEVNHACLPOAAACAPDLPAFSGGF 759

```

RESULT 4 ACOX_BOVIN

```

ID ACOX_BOVIN STANDARD: PRT: 762 AA.
AC 029457
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Copper amine oxidase, liver isozyme precursor (EC 1.4.3.6) (Amine
OS oxidase [copper-containing]) (Serum amine oxidase) (SAO).
OC Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
RN (1)
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RA MEDLINE=94193686; PubMed=8144587;
RA Mu D., Medzhirsdzky K.F., Adams G.W., Mayer P., Hines W.M.,
RA Burlingame A.L., Smith A.J., Cal D., Kilman J.P.;
RT "Primary structures for a mammalian cellular and serum copper amine
RT oxidase."
RL J. Biol. Chem. 269:9926-9932(1994).

```

```

RN (2)
RP PARTIAL SEQUENCE.
RX MEDLINE=93090748; PubMed=1457410;
RA Jones S.M., Palcic M.M., Scaman C.H., Smith A.J., Brown D.E.,
RA Dooley D.M., Mure M., Kilman J.P.;
RT Identification of topaquinone and its consensus sequence in copper
RT amine oxidases."
RL Biochemistry 31:12147-12154(1992).
CC 1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
CC H(2)O(2).
CC 1- COFACTOR: Binds 1 copper ion and 1 topaquinone per subunit.
CC 1- ENZYME REGULATION: INHIBITED BY AMILORIDE IN A COMPETITIVE MANNER.
CC 1- SUBCELLULAR LOCATION: Extracellular.
CC 1- TISSUE SPECIFICITY: Liver.
CC 1- PTM: Topaquinone (TPQ) is generated by copper-dependent
CC autoxidation of a specific tyrosyl residue (By similarity).
CC 1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: S69583; AAB30397.1; -.
CC EMBL: L27218; AAA30525.1; -.
CC InterPro: IPR000269; CUNH_oxidase.
CC Pfam: PF01179; Cu_amine_oxid_1.
CC Pfam: PF02727; Cu_amine_oxid_2; 1.
CC Pfam: PF02728; Cu_amine_oxid_3; 1.
CC PRINTS: PR00766; CUDAOXIDASE.
CC PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
CC PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
CC KMOxidoreductase; Copper; TPQ; Glycoprotein; signal; Metal-binding.
CC
CC FT CHAIN 1 762 COPPER AMINE OXIDASE, LIVER ISOZYME.
CC FT MOD_RES 470 470 TOPAQUINONE (By SIMILARITY).
CC FT METAL 519 519 COPPER (POTENTIAL).
CC FT METAL 521 521 COPPER (POTENTIAL).
CC FT METAL 683 683 AMILORIDE (By SIMILARITY).
CC FT BINDING 672 672 AMILORIDE (By SIMILARITY).
CC FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 665 665 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 762 AA; 84756 MW; AA959771360295FE CRC64;

```

Query Match 81.4%; Score 3334; DB 1; Length 762;
Best Local Similarity 80.6%; Pred. No. 3; le-252;
Matches 610; Conservative 61; Mismatches 76; Indels 10; Gaps 2;

```

OY 14 VITFALVYLAVGR--GDDGE-----PSQLPHCPSPVSPSAQPTWTRPGSQSLFADL 63
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 3 IFIFLSLMTLLVNGREBGGVSEGVGKCHPSLPRCPSPSPSDQPTWTRPGSQSLFADL 62
OY 64 SRELTAVMFLTORLGRGLVDAQAARPSDNCVSEVLOLPPKAAALAHLDKSSPPARE 123
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 63 SRELTAVMFLTORLGRGLVDAQAARPSDNCVSEVLOLPPKAAALAHLDKSSPPARE 122
OY 124 ALAIVFGROPORNVSELVVGPLPHPSYMDVYVERHGGLPYRRRPVLFQEYLDIDOMI 183
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 123 ALAIVFGROPORNVSELVVGPLPHPSYMDVYVERHGGLPYRRRPVLFQEYLDIDOMI 182
OY 184 FNNELPOASGLLHHCCEYKRGKRLVMTTAPRGLSGDRATWFGLYYNSGAGFLAHV 243
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 183 FNNELPOAAGVLLHHCCEYKRGKRLVMTTAPRGLSGDRATWFGLYYNSGAGFLAHV 242
OY 244 GLELLVNHKALDPAKMTIOKVFYQGRYDSLAOLEAFGLVNVVLIIPNGTGSMSLK 303
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 243 GLELLVNHKALDPAKMTIOKVFYQGRYDSLAOLEAFGLVNVVLIIPNGTGSMSLK 302
OY 304 SPVPGPAPPLQFYPOGPRFSVQGSRAVSSSLMTFSEGLGAFSGPRIPDFVQERLYEYI 363

```

```

Db 303 SOVPEPTPLPLOGHPOGPFPSVGNVASSLMTFSGLGFSGPRFVDVAFQGERLAYET 362
Oy 364 SLOEALAIYGNSPAAMATTRYVDGFGMGKRYTTLTRGVDCPYLATYVDMHFLLESQAPK 423
Db 363 SLOEAGAYVGNTPAAMLTRFYMDSGFMGFATPLIRGVDCPYLATYVDMHFLVESQTPK 422
Oy 424 TIDAFCEVEQNOGLPLRRHSDLYSHYRGGLAEVLYVRSKSTLNTDYMDTVFHPSG 483
Db 423 TLHDACVEQONKGLPLRRHSDLYSHYRGGLAEVLYVRSKSTLNTDYMDTVFHPSG 482
Oy 484 AIRIRYATGYISSAFLEFGATGYGNVSEHTGTHTSHAHKRVLDVAGLEWMAED 543
Db 483 AIEVKLHAATGYISSAFLEFGAARRKGNVGEHTLGPVHTSHAHKRVLDVAGLEWMAED 542
Oy 544 MVEFPAVWSPBHQRLQVTRKRLLEMEBOAFLVGSATPRYLYLASNHNKMGHPRGY 603
Db 543 MAVPFAIPWSPBHQRLQVTRKRLLEMEBOAFLVGSATPRYLYLASNHNKMGHPRGY 602
Oy 604 RIOMLSEFAGEPLPQNSSMARGFSEWERYQLAVTORKEBESSSVFNQNDPMATTVDSDF 663
Db 603 RIQTVSEFAGPMPQNSPMEAFSGRQLAITORKETEPSSSVFNQNDPMATTVDSDF 662
Oy 664 INNETIAGKDLVAMTAGFLPHIHAEDIPMTVTVGVNGVGFELRPYNEFDEDPSPYSADSI 723
Db 663 INNETIAGKDLVAMTAGFLPHIHAEDIPMTVTVGVNGVGFELRPYNEFDEDPSPYSADSI 722
Oy 724 YFREGDODAGACEVNPPLACLPQAACAAPDLPAFSHGGE 760
Db 723 YFREGDODAGACEVNPPLACLPQAACAAPDLPAFSHGGE 759

RESULT 5
AOC2_HUMAN
ID AOC2_HUMAN STANDARD; PRT; 756 AA.
AC 075106; 075105; 000120; 09UNY0;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Retina-specific copper amine oxidase precursor (EC 1.4.3.6) (RAO)
DE (Amine oxidase [copper-containing]).
GN AOC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=97237047; PubMed=9119395;
RA Imanura Y., Kubota R., Wang Y., Asakawa S., Kudoh J., Mashima Y.,
RA Oguchi Y., Shimizu N.;
RT "Human retina-specific amine oxidase (RAO): cDNA cloning, tissue
RT expression, and chromosomal mapping.";
RL Genomics 40:277-283(1997).
RN [2]
RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RX MEDLINE=98390194; PubMed=9722934;
RA Imanura Y., Noda S., Mashima Y., Kudoh J., Oguchi Y., Shimizu N.,
RT "Human retina-specific amine oxidase: genomic structure of the gene
RT (AOC2), alternatively spliced variant, and mRNA expression in
RT retina.";
RL Genomics 51:293-298(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RA Zhang X., McIntire W.S.;
RT "Human copper-containing amine oxidases";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY BE A CRITICAL MODULATOR OF SIGNAL TRANSMISSION IN
CC RETINA, POSSIBLY BY DEGRADING THE BIOGENIC AMINES DOPAMINE,
CC HISTAMINE, AND PUTRESCINE.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +

```

```

CC H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
CC HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: RETINAL-SPECIFIC.
CC -1- PTM: Topaquinine (TPQ) is generated by copper-dependent
CC autoxidation of a specific tyrosyl residue (By similarity).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC
DR EMBL: AB012943; BAA32590.1; -.
DR EMBL: AB012942; BAA32590.1; JOINED.
DR EMBL: AB012943; BAA32589.1; -.
DR EMBL: AB012942; BAA32589.1; JOINED.
DR EMBL: D88213; BAA19001.1; -.
DR EMBL: AF081363; AAD33345.1; -.
DR Genew: HGNC:549; AOC2.
DR MIM: 602268; -.
DR InterPro: IPR000269; CuNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid_1.
DR Pfam: PF02727; Cu_amine_oxidn2; 1.
DR Pfam: PF02728; Cu_amine_oxidn3; 1.
DR PRINTS: PR00766; CUADOXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
DR Oxidoreductase; Copper; TPQ; Alternative splicing; Glycoprotein;
KW Signal; Metal-binding.
KM
FT SIGNAL 1 32
FT CHAIN 33 756
FT MOD_RES 465 465
FT METAL 516 516
FT METAL 518 518
FT METAL 680 680
FT BINDING 669 669
FT CARBOHYD 133 133
FT CARBOHYD 198 198
FT CARBOHYD 226 226
FT CARBOHYD 388 388
FT CARBOHYD 662 662
FT VARSPLIC 599 625
FT CONFLICT 181 181
FT CONFLICT 215 218
FT CONFLICT 221 222
FT CONFLICT 610 610
SQ SEQUENCE 756 AA; 83793 MW; 4E3B7317ED6DCA66 CRC64;

Query Match 63.8%; Score 2612; DB 1; Length 756;
Best Local Similarity 65.0%; Pred. No. 6; 6e-196;
Matches 495; Conservative 92; Mismatches 167; Indels 8; Gaps 3;

Oy 1 MNOKTILLVLLAVITLALVCLVGVGGDGEQSLPHCPSPVSAQPFTHGQSQLF 60
Db 1 MHLKIYAFALSLITLIFALAYVLLTSPGS----SQPHCPVSHRAQWPHHGQSQLF 56
Oy 61 ADSRELLTAVMFLTQRLPGVLDAQAQARPNDNCVSVELQLPKKAALAHDRGSP 120
Db 57 ADSRELLTAVMFLTQRLPGVLDAQAQARPNDNCVSVELQLPKKAALAHDRGSP 116
Oy 121 AREALAVFGRPOPVNSVLYVGPPLPSPYMRDVTERRGGPLPYRRPVLFOEYDID 180
Db 117 AREALAVFEGGQPOPVNSVLYVGPPLPSPYMRDVTERRGGPLPYRRPVLFAEFQW 176
Oy 181 QMIFNRELPOASGLLHHCCEYKRRGNLVTMTAPRGLQSGDRATWFGLYNIGAGFFL 240
Db 177 RHLDKVELPKAPFLSLST--FNNGSTLAAVHAHPRLRSRERTYMAALVHNISGVCLF 234

```

```

QY 241 HHVGLLELVNKKALDPARMTIQKVFYOGRYVDSIAOLEAGFLVNVLPIDNGTGSW 300
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DQ 235 HPEVLELLDHRALDPAMHTVQGVFYGLHYADLGQLEREKSRGLVRRPLPPNGAS 294
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 301 SLKSPVPPGAPPLQGYPOGPRSVGSSRAVSSLSMTSFGISGAPRIEDVREGERLV 360
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DQ 295 SLRSRNSPFGPLPQFSPOGSOYVSGCNLVVSSLSMFFGCHGVSGRIFFDROGERIA 354
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 361 YEISLOALIVYGNNSPAAAMTRVVDGFGMKYTTPLTRVDQPYLATYVDHMFLESO 420
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DQ 355 YEVSQGVSVSYGADSKRYMTLRILDSDFGLGRNSRGLVRCVDPQATVWDIHLVKGK 414
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 421 APTKIRDAFCVEEQNOGLPLRRHSDLSHYFGSLAEVLVVRSMSTLNDYVWDVFEH 480
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DQ 415 AVQLLPBACVCFEEAAGLPLRRHNYLONHRYGGLSALVYRVSSTGNDIYIDFELY 474
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 481 PSGAIEIRFATGTITSSAFLEGATG--KYNQVSEHTLGVTHSAHEKVDLVAAGLENN 538
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DQ 475 PNGALEGRVHATGYINTAFILKGEGELFGNRVGEERYLGVTHAHPKILDLVAAGLKNW 534
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 539 VMAEDVFPVAPVSPPEHOLQRLQVTRKILMEEOAFLVGSATPRYLYLASHSKMG 598
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DQ 535 VVAEDVFKFVAAPWNEHMLQRPQLTRQVKGEDLTFASGLPLRYLYLASHTQTNMG 594
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 599 HPRGRTQMTSFAGEPLPONSNAFGSMERYQLAVTQKEEPESSSSVFQNDPMAPT 658
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DQ 595 HORGRTQIHSPLGIQPLSDMERALSMGRYQLVTVQKEEEOSSSTIYQNDIMPTV 654
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 659 DESFINNETIAGKDYAAWYTAGFLHPHAEDIPNTVTVGNGVFELRPYVFEDEPSEY 718
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DQ 655 TFADFNNETLIGEDLYAAWYTAGFLHPHAEDIPNTVTVGNGVFELRPYVFEDEPSEY 714
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 719 SADSIFYRGDODAGACEVNPACLPQAAACAPDLPAFSGHGF 760
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DQ 715 SPGSVYFEKGODAGLCSINPACLPDLAACPDLPPSYHGF 756
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :

```

```

CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinone per subunit.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
CC -1- PTM: Topaquinone (TPO) is generated by copper-dependent
CC autoxidation of a specific tyrosyl residue (By similarity).
CC -1- PTM: N- AND O-GLYCOSYLATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPOQUINONE OXIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC
DR EMBL: U72632; AAC53189.1; -.
DR InterPro: IPR000269; CuNH_oxidase.
DR Pfam: PF02727; Cu_amine_oxid2; 1.
DR Pfam: PF02728; Cu_amine_oxid3; 1.
DR PRINTS: PR00766; CUDAOXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; PARTIAL.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; PARTIAL.
DR Oxidoreductase; Copper; TPO; Glycoprotein; Transmembrane;
KW Signal-anchor; Cell adhesion; Metal-binding.
FT INT_MET 0 0
FT DOMAIN 1 5 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 6 26 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 27 >321 EXTRACELLULAR (POTENTIAL).
FT FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 321 321
SQ SEQUENCE 321 AA; 35116 MW; 6B1F294253A1DE6F CRC64;

Query Match 34.6%; Score 1418; DB 1; Length 321;
Best Local Similarity 83.4%; Pred. No. 2,9e-103;
Matches 267; Conservative 18; Mismatches 35; Indels 0; Gaps 0;

QY 3 OKTLLVLLILAVITIFALVYLVLRGDDGSPQLPHCPSPVSPAPMTWPGSQSLFAD 62
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DQ 2 OKTLLVLLILAVITIFALVYLVLRGDDGSPQLPHCPSPVSPAPMTWPGSQSLFAD 61
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 63 LSREITAVMFELOKRGVLDAAQAPSDNCVSVELQAPRAALAHNRDSSPPAR 122
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DQ 62 LSPELTVAMSFLLIKHGLGVLDAQAAPSDNCVSVELQAPRAALAHNRDSSPPAR 121
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 123 EALAVFGRPOPVNSVSLVGPLPHPSYMDYVERHGGPLPYRRPVLYFOEYLDIDOM 182
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DQ 122 EALAVFGRPOPVNSVSLVGPLPHPSYMDYVERHGGPLPYRRPVLYFOEYLDIDOM 181
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 183 IFNRELPOASGLLHHCFFYKHKGRNLYVTMTAPRGLOGSDRATVFGLYNISGAGFLAH 242
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DQ 182 IFNRELPOASGLLHHCFFYKHKGRNLYVTMTAPRGLOGSDRATVFGLYNISGAGFLAH 241
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 243 VGLLELVNKKALDPARMTIQKVFYOGRYVDSIAOLEAGFLVNVLPIDNGTGSW 302
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DQ 242 VGLLELVNKKALDPARMTIQKVFYOGRYVDSIAOLEAGFLVNVLPIDNGTGSW 301
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
QY 303 KSPVPPGAPPLQGYPOGPR 322
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DQ 302 KSPVPPGAPPLQGYPOGPR 321
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :

```

```

RESULT 7
ABP_RAT
ID ABP_RAT STANDARD; PRT; 746 AA.
AC P36633;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Amiloride-sensitive amine oxidase [copper-containing] precursor

```

Query Match	34.2%	Score 1400.5;	DB 1;	Length 746;
Best Local Similarity	-39.1%;	Pred. No. 2.2e-101;		
Matches 300;	Conservative 131;	Mismatches 280;	Indels 57;	Gaps 14;

D	b	51	REELGLDQPSKEPLTANKSVLEIMLPKKKHVLKFLDEGRKQPNREARAVIFFGQADY	110
Q	y	138	VSELYVGRPLPHPSYMDVYVERNGGLP ¹ PHRRPVLEQ ² VELDDIDQ ³ MFNELQ ⁴ ASGLLH	197
D	b	111	VTEFVAGPLPRPYIRL ⁵ S-PRGNHLSWSRPISTAE ⁶ ---DLLYH-TLKRATP ⁷ MLQ	164
Q	y	198	-----CCEYKRGRLVMTAPRGLSGSDATVFG ⁸ LYTNISGAGFELHVGLELVN	250
D	b	165	FLDDTGTGFSFLGDDDLT ⁹ TDVAPRGVASGQSRMFIQRYVE-GYLLHPTGLEILLD	222
Q	y	251	HKALDPARKWTIQVYQGRYDLSLAOLEAFEGAGLVNVLID ¹⁰ --NGTGSWSLKS ¹¹ SPV	306
D	b	223	HGSTVDYDMVNEEDLWNGK ¹² YNNPEELARKYAA ¹³ GEVDVJVLDEPLRPNGTE---KPL ¹⁴ LE	277
Q	y	307	-----PPGAPRLD ¹⁵ FOYPOGREFSGSRV ¹⁶ ASLSLMTESFGCAGSGRIP ¹⁷ VRQ	355
D	b	278	SSYKPRGEFT ¹⁸ PPVNAAGPHVOPSGRRKIEGNTVLYGKMSYSYLRSSGLQIT ¹⁹ NVLF	337
Q	y	356	GERLVEISLQELALAYGNSPAA ²⁰ MTT ²¹ RYVDGGFGMKY ²² TPLTRGVD ²³ PLATYVD ²⁴ MHF	415
D	b	338	GERVAYEVSVQEAVALYGH ²⁵ T ²⁶ PAGMOTKYIDVGMGLSV ²⁷ THELAPGIDC ²⁸ PE ²⁹ TAF ³⁰ LD ³¹ AFH	397
Q	y	416	LLSEQA ³² K ³³ IT ³⁴ RD ³⁵ AFVCEQ ³⁶ QGL ³⁷ PLRRHSDLS---HFGGIAEVL ³⁸ LV ³⁹ BSM ⁴⁰ STLND	472
D	b	398	YVSDGGVNHYPHALC ⁴¹ LEFEM ⁴² TPGLRRHFN ⁴³ SNKGFEN ⁴⁴ YAGLKG ⁴⁵ VL ⁴⁶ LR ⁴⁷ TS ⁴⁸ TVYND	457
Q	y	473	YVMDTVPHPGALIEIR ⁴⁹ FYAYGISAFLEF ⁵⁰ GATK ⁵¹ YGNQ ⁵² SEHT ⁵³ LG ⁵⁴ VH ⁵⁵ SAH ⁵⁶ FKV ⁵⁷ LDY	532
D	b	458	YIMDFIYSNGV ⁵⁸ EA ⁵⁹ KH ⁶⁰ AT ⁶¹ GVYHAF ⁶² Y ⁶³ PEGLRHGT ⁶⁴ RLQ ⁶⁵ THLGN ⁶⁶ TH ⁶⁷ LVH ⁶⁸ YRDMY	517
Q	y	533	AGLENNYMAEDMYFVPR ⁶⁹ AVWSP ⁷⁰ EPHOLQ ⁷¹ RLQ ⁷² TRKLL ⁷³ MEBQ ⁷⁴ AAFLV ⁷⁵ GSAT ⁷⁶ TR ⁷⁷ Y ⁷⁸ VLASN	592
D	b	518	AGTNS ⁷⁹ FQ ⁸⁰ LT ⁸¹ MLK ⁸² LENT ⁸³ IN ⁸⁴ QWSP ⁸⁵ SHSLY ⁸⁶ PT ⁸⁷ LEQ ⁸⁸ Y ⁸⁹ SOE ⁹⁰ HO ⁹¹ AA ⁹² FF ⁹³ Q ⁹⁴ LT ⁹⁵ LEK ⁹⁶ YL ⁹⁷ LFSSP	577
Q	y	593	HSNKGHPGRYRIOMLS ⁹⁸ FACE ⁹⁹ PLPONS ¹⁰⁰ SMAR ¹⁰¹ GF ¹⁰² SMERYOLAV ¹⁰³ ORKEE ¹⁰⁴ PS ¹⁰⁵ SSV ¹⁰⁶ ENOMD	652
D	b	578	QKNQMGHRR ¹⁰⁷ REY ¹⁰⁸ RLQ ¹⁰⁹ ISH ¹¹⁰ MAQV ¹¹¹ LPCKQ ¹¹² GE ¹¹³ RAV ¹¹⁴ TAN ¹¹⁵ RLY ¹¹⁶ PLAT ¹¹⁷ Y ¹¹⁸ KY ¹¹⁹ ES ¹²⁰ ERY ¹²¹ SSSL ¹²² NOND	637
Q	y	653	PWATVDP ¹²³ SD ¹²⁴ FT-NNETI ¹²⁵ AK ¹²⁶ DLV ¹²⁷ AVY ¹²⁸ TAG ¹²⁹ FL ¹³⁰ HPA ¹³¹ EDI ¹³² PN ¹³³ TV ¹³⁴ GV ¹³⁵ NGV ¹³⁶ FF ¹³⁷ LP ¹³⁸ RYNEF	711
D	b	638	PMP ¹³⁹ PPV ¹⁴⁰ YEE ¹⁴¹ FL ¹⁴² RNN ¹⁴³ ENIED ¹⁴⁴ ELV ¹⁴⁵ AWY ¹⁴⁶ GF ¹⁴⁷ LHP ¹⁴⁸ ISE ¹⁴⁹ DV ¹⁵⁰ PN ¹⁵¹ AT ¹⁵² GC ¹⁵³ NSV ¹⁵⁴ GL ¹⁵⁵ LRP ¹⁵⁶ NF	697
Q	y	712	DEDSFYSADSI ¹⁵⁷ FRGDOD ¹⁵⁸ GAC ¹⁵⁹ EVN ¹⁶⁰ PL ¹⁶¹ QD ¹⁶² PAAC ¹⁶³ AD ¹⁶⁴ LP ¹⁶⁵ AF ¹⁶⁶ SHG	759
D	b	698	PEDSLAS ¹⁶⁷ RDT ¹⁶⁸ VIV-WPQD ¹⁶⁹ GLNR ¹⁷⁰ VO--RWIP ¹⁷¹ DRCL ¹⁷² VS-PPFS ¹⁷³ YNG	741
RESULT 8				
ABP_HUMAN				
ID	ABP_HUMAN	STANDARD:	PRT:	751 AA.
AC	P19801; Q16683; Q16684;			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Amloride-sensitive amine oxidase [copper-containing] precursor (EC 1.4.3.6) (diamine oxidase) (DAO) (Amloride-binding protein) (ABP) (Histaminase).			
DE	ABP1 OR DAO1 OR AOCL1.			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Choriata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_Taxid=9606;			
RN	11}			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94237856; PubMed=8182053;			
RA	Chassande O., Renard S., Barbry P., Iazdunski M.;			
RT	"The human gene for diamine oxidase, an amiloride binding protein.			
RT	Molecular cloning, sequencing, and characterization of the promoter."			
RL	J. Biol. Chem. 269:14484-14489(1994).			
RN	12}			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Placenta;			
RX	MEDLINE=96113540; PubMed=8595053;			

RA Zhang X., Kim J., McIntire W.S.;
 RT "cDNA sequences of variant forms of human placenta diamine oxidase";
 RL Biochem. Genet. 33:261-268(1995).
 [3]
 RN PRELIMINARY SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RP TISSUE-KIDNEY;
 RC MEDLINE-91017502; PubMed-2217167;
 RX Barbry P., Champe M., Chassande O., Munemitsu S., Champigny G.,
 RA Lingneglia E., Maes P., Frelin C., Tarter A., Ullrich A.,
 RA Lazdunski M.,
 RT "Human kidney amiloride-binding protein: cDNA structure and functional
 expression";
 RT Proc. Natl. Acad. Sci. U.S.A. 87:7347-7351(1990).
 [4]
 RN SEQUENCE OF 20-39, AND CHARACTERIZATION.
 RP TISSUE-Placenta;
 RX MEDLINE-94193685; PubMed-8144586;
 RA Novotny W.F., Chassande O., Baker M., Lazdunski M., Barbry P.;
 RT "Diamine oxidase is the amiloride-binding protein and is inhibited by
 amiloride analogues";
 RL J. Biol. Chem. 269:9921-9925(1994).
 CC -1- FUNCTION: CATALYZES THE DEGRADATION OF COMPOUNDS SUCH AS
 PUTRESCINE, HISTAMINE, SPERMINE, AND SPERIDINE, SUBSTANCES
 INVOLVED IN ALLERGIC AND IMMUNE RESPONSES, CELL PROLIFERATION,
 TISSUE DIFFERENTIATION, TUMOR FORMATION, AND POSSIBLY APOPTOSIS.
 CC PLACENTAL DAO IS THOUGHT TO PLAY A ROLE IN THE REGULATION OF THE
 CC FEMALE REPRODUCTIVE FUNCTION.
 CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
 H(2)O(2).
 CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
 CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: PLACENTA AND KIDNEY.
 CC -1- PTM: Topaquinine (TPQ) is generated by copper-dependent
 autoxidation of a specific tyrosyl residue.
 CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X78212; CA55046.1; -;
 DR EMBL: U11863; AAC50270.1; -;
 DR EMBL: U11863; AAB60381.1; -;
 DR EMBL: M55602; AAA58358.1; ALT_SEQ.
 DR PIR: A38276; A38276.
 DR PIR: S42495; S42495.
 DR Genew: HGNC:80; ABP1.
 DR MIM: 104610; -;
 DR InterPro: IPR000269; CuNH_oxidase.
 DR Pfam: PF01179; Cu_amine_oxid; 1.
 DR Pfam: PF02727; Cu_amine_oxidn3; 1.
 DR Pfam: PF02728; Cu_amine_oxidn3; 1.
 DR PRINTS: PR00766; CUDAOXIDASE.
 DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
 DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
 DR Signal: Glycoprotein; Oxidoreductase; Copper; Heparin-binding; TPQ;
 KW Alternative splicing; Polymorphism; Metal-binding.
 FT SIGNAL 1
 FT CHAIN 19
 FT METAL 391
 FT MOD_RES 461
 FT METAL 510
 FT METAL 512
 FT METAL 675
 FT BINDING 568
 FT BINDING 664
 FT CARBOHYD 110

FT CARBOHYD 168 168 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 538 538 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 745 745 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARSPLIC 619 619 R -> RTEGQPALSOARSPVGR (IN ISOFORM
 FT VARIANT 645 645 DAO2).
 FT VARIANT 645 645 H -> D.
 FT CONFLICT 28 28 /FTID=VAR_007542.
 FT CONFLICT 332 332 R -> A (IN REF. 2).
 FT CONFLICT 332 332 S -> F (IN REF. 2).
 SQ SEQUENCE 751 AA; 85363 MW; 1B8B695C44232DA CRC64;
 Query Match 34.0%; Score 1394; DB 1; Length 751;
 Best Local Similarity 39.8%; Pred. No. 7.1e-101;
 Matches 293; Conservative 124; Mismatches 281; Indels 38; Gaps 11;
 50 PWRPGSOLFADLSREBLVAMRFLTORLGPGLVDAQAARSPNCVFSVLOLPPRAAA 109
 23 PGLTPRAGVSDLSNOELKAVHSFLWSKELRLQPSSTTMANTVLEMLPKKYHV 82
 110 LAHLDRGPPRARELAIVFGRPOPVSSELYVGPPLPHEPMYMDVYERHGGFLPHRR 169
 83 LRFIDGRHFRARAVIFGDEGHPVTEFAVGPLPGCYMALSPRGYSSMASR 141
 170 PVLFOEYLDIDOMTFNRELPOASGLNH-----CCFYKHGRNLYMTTAPRGLOSGD 222
 142 PISTAEY-----ALLYHTLOEATKPHQFLNTGFSFQDCHRCCLAFDVAARGVASQ 196
 223 RATHFGLYNSGAGFLHNGLELLVNHKALDPAKRTQKVFQGRYDSLAOLEQFE 282
 197 RRSMLIIORYVE-GYFLHPTGELLELDVHGSDGSHNAVEQVWYNGKFGSPLEAKRYA 254
 283 AGLVNVLIPNGTGGSSLSKSPVP-----GPAP-----PLQFYQGRFSGV 327
 255 DGEVDVVLLEPLDPLGGKCHDSTEERPLFSHKRPGDPPSPHVGPRVLDVPHGRFRLEG 314
 328 SRVASSLWTFSGFGLAEGSPRIEDVRFQGERLYVEISLQELALAYGNSPAAATTTRYVDG 387
 315 NAVLYGGSFARLRSSGGLQVLNVHGEERIAVEVSQEAVALYGGHTPAGMGTXYDV 374
 388 GFGGKGYTPPLTRGVDCYLTATYDMHLESQAPKTRQAFVFEQNOGLPLRRHNSDL 447
 375 GWSGLSVTHLAPIDCEPTATFLDTYHYDADPVHYPRALCFEMPTGVPLRRHNSN 434
 448 YS---HYEGGLAETVLVVRSMSTLNDYVDYVDFHSGAIEIFYATGYSSAFLGAT 504
 435 FKGGFNFYAGLKGVLVLRITSTYVNDYINDYIFEPNGVEMAHMNTGYHAATFYPRG 494
 505 GRYGNVSEHTLGTVHTSHAHFKYDLDVAGLENNVMAEDVFEVMAVPMSPHOLQRLQV 564
 495 LRHGTRLTHTLIGHIHTHLVYRVDLDVAGTKNSFOTLQMKLENTITNPMSPRHHVVOPTL 554
 565 TRKILMEEOAFLVGSATPRYLVLASHSNKGHPRGYRIOMLSFAGEPLPQSSMAR 624
 555 EOTOYSWEKQAFRRKRLPKYLLFTSPQENPMGCHKRSYRLQIHSMDQVLPQMOEBQ 614
 625 FSWERYQLAVYQKEEPESSSSVNONDPAPATYDFSDFI--NNETISGKDLVAVATGFL 683
 615 ITMARPLAVTKYSESLSCSSSIYHONDPMPRVYEDQLNNENIEDLVAVATYVGL 674
 684 HIPHAEDIPNTVNGVNGVGFPLRPYNFEDEPSTYSADSIYFRQDQAGACVNPPLACL 743
 675 HIPSEDPNATPGNSVGFLLRPENFEDEPDSLASRDTYIV-WPRONGPMYVQ--RWIP 731
 744 QAAACAPDLPAFSGHG 759
 732 EDRDCSMP-PPFSYNG 746
 RESULT 9
 AMO_KLEAE STANDARD; PRT; 755 AA.
 ID AMO_KLEAE
 AC P49250;
 DT 01-FEB-1996 (Rel. 33, Created)

```

DT 01-FEB-1996 (Rel. 33, last sequence update)
DE 15-JUN-2002 (Rel. 41, last annotation update)
DE Copper amine oxidase precursor (EC 1.4.3.6) (Monamine oxidase)
DE (Tyramine oxidase).
GN MAOA OR TYNA.
OS Klebsiella aerogenes.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxId=28451;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-41.
RC STRAIN=M70;
RX MEDLINE=92210491; PubMed=1556068;
RA Sugino H., Sasaki M., Azakami H., Yamashita M., Murooka Y.;
RT "A monamine-regulated Klebsiella aerogenes operon containing the
RL monamine oxidase structural gene (maoA) and the maoc gene.";
J. Bacteriol. 174:2485-2492(1992).
CC -1- FUNCTION: ACTIVE ON TYRAMINE, TRYPTAMINE, BETA-PHENETHYLAMINE AND
CC DOPAMINE.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
CC H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaguinone per subunit.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- INDUCTION: BY TYRAMINE AND CATECHOLAMINES.
CC -1- PTM: Topaguinone (TPQ) is generated by copper-dependent
CC autoxidation of a specific tyrosyl residue (by similarity).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAGUINONE OXIDASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10208; BAA01060.1; -.
DR HSSP: P46883; IOAC.
DR InterPro: IPR00269; CuNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid_1.
DR Pfam: PF02727; Cu_amine_oxid2; 1.
DR Pfam: PF02728; Cu_amine_oxid3; 1.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
DR OXidoreductase; Copper; TPQ; Periplasmic; Signal; Metal-binding.
FT SIGNAL 1 30
FT CHAIN 31 755 COPPER AMINE OXIDASE.
FT ACT SITE 413 413 CATALYTIC BASE (BY SIMILARITY).
FT MOD RES 496 496 TOPAGUINONE (BY SIMILARITY).
FT METAL 554 554 COPPER (POTENTIAL).
FT METAL 556 556 COPPER (POTENTIAL).
FT METAL 719 719 COPPER (POTENTIAL).
SQ SEQUENCE 755 AA; 83576 MW; 7B5552283CD93EFC CRC64;

Query Match 10.1%; Score 413; DB 1; Length 755;
Best Local Similarity 27.5%; Pred. No. 2.5e-24;
Matches 150; Conservative 75; Mismatches 213; Indels 108; Gaps 21;

```

```

OY 442 RHHSGLYSKFFG-----LAETLVYVRSMSTLNDYVDVTPHPSGATIRFYATGYIS 486
DE 15-JUN-2002 (Rel. 41, last annotation update)
DE Copper amine oxidase precursor (EC 1.4.3.6) (Monamine oxidase)
DE (Tyramine oxidase).
GN MAOA OR TYNA.
OS Klebsiella aerogenes.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxId=28451;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-41.
RC STRAIN=M70;
RX MEDLINE=92210491; PubMed=1556068;
RA Sugino H., Sasaki M., Azakami H., Yamashita M., Murooka Y.;
RT "A monamine-regulated Klebsiella aerogenes operon containing the
RL monamine oxidase structural gene (maoA) and the maoc gene.";
J. Bacteriol. 174:2485-2492(1992).
CC -1- FUNCTION: ACTIVE ON TYRAMINE, TRYPTAMINE, BETA-PHENETHYLAMINE AND
CC DOPAMINE.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
CC H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaguinone per subunit.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- INDUCTION: BY TYRAMINE AND CATECHOLAMINES.
CC -1- PTM: Topaguinone (TPQ) is generated by copper-dependent
CC autoxidation of a specific tyrosyl residue (by similarity).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAGUINONE OXIDASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10208; BAA01060.1; -.
DR HSSP: P46883; IOAC.
DR InterPro: IPR00269; CuNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid_1.
DR Pfam: PF02727; Cu_amine_oxid2; 1.
DR Pfam: PF02728; Cu_amine_oxid3; 1.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
DR OXidoreductase; Copper; TPQ; Periplasmic; Signal; Metal-binding.
FT SIGNAL 1 30
FT CHAIN 31 755 COPPER AMINE OXIDASE.
FT ACT SITE 413 413 CATALYTIC BASE (BY SIMILARITY).
FT MOD RES 496 496 TOPAGUINONE (BY SIMILARITY).
FT METAL 554 554 COPPER (POTENTIAL).
FT METAL 556 556 COPPER (POTENTIAL).
FT METAL 719 719 COPPER (POTENTIAL).
SQ SEQUENCE 755 AA; 83576 MW; 7B5552283CD93EFC CRC64;

Query Match 10.1%; Score 413; DB 1; Length 755;
Best Local Similarity 27.5%; Pred. No. 2.5e-24;
Matches 150; Conservative 75; Mismatches 213; Indels 108; Gaps 21;

```


RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
 RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
 RA Oshima T., Saito N., Sampel G., Seki Y., Sivasondaram S.,
 RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Mada C.,
 RA Yamamoto Y., Horiuchi T.,
 RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 RN [5]
 RP SEQUENCE OF 1-59 FROM N.A.
 RC STRAIN-W / ATCC 11105;
 RX MEDLINE=98421522; PubMed=9748275;
 RA Fernandez A., Mnamdres B., Garcia B., Olivera E.R., Luengo J.M.,
 RA Garcia J.L., Diaz E.;
 RT "Catabolism of phenylacetic acid in *Escherichia coli*. Characterization
 RT of a new aerobic hybrid pathway.";
 RL J. Biol. Chem. 273:25974-25986(1998).
 RN [6]
 RP SEQUENCE OF 477-757 FROM N.A.
 RC STRAIN-W / ATCC 11105;
 RX MEDLINE=97263463; PubMed=9109378;
 RA Fernandez A., Prieto M.A., Garcia J.L., Diaz E.;
 RT "Molecular characterization of *padA*, a phenylacetaldehyde
 RT dehydrogenase from *Escherichia coli*.";
 RL FEBS Lett. 406:23-27(1997).
 RN [7]
 RP SEQUENCE OF 31-50.
 RX MEDLINE=96213037; PubMed=8631685;
 RA Yamagishi M., Azakami H., Yokoro N., Roh J.-H., Suzuki H.,
 RA Kumagai H., Murooka Y.;
 RT "mobA, a gene that encodes a positive regulator of the monamine
 RT oxidase gene (*moaA*) in *Escherichia coli*.";
 RL J. Bacteriol. 178:2941-2947(1996).
 RN [8]
 RP SEQUENCE OF 31-40.
 RX STRAIN-K12;
 RX MEDLINE=97195795; PubMed=9043126;
 RA Hanlon S.P., Hill T.K., Flavell M.A., Stringfellow J.M., Cooper R.A.;
 RT "2-phenylethylamine catabolism by *Escherichia coli* K-12: gene
 RT organization and expression.";
 RL Microbiology 143:513-518(1997).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=97200715; PubMed=9048544;
 RA Wilmot C.M., Murray J.M., Alton G., Parsons M.R., Convery M.A.,
 RA Blakeley V., Corner J.A.S., Palcic M.M., Knowles P.F., McPherson M.J.,
 RA Phillips S.E.V.;
 RT "Catalytic mechanism of the guinoenzyme amine oxidase from
 RT *Escherichia coli*: exploring the reductive half-reaction.";
 RL Biochemistry 36:1608-1620(1997).
 RN [10]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=99315198; PubMed=10387067;
 RA Murray J.M., Szebell C.G., Wilmot C.M., Tambyrajah W.S., Jaeger J.,
 RA Knowles P.F., Phillips S.E.V., McPherson M.J.;
 RT "The active site base controls cofactor reactivity in *Escherichia*
 RT *coli* amine oxidase: X-ray crystallographic studies with mutational
 RT variants.";
 RL Biochemistry 38:8217-8227(1999).
 RN [11]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
 RX MEDLINE=20045000; PubMed=10576737;
 RA Wilmot C.M., Hajdu J., McPherson M.J., Knowles P.F., Phillips S.E.V.;
 RT "Visualisation of dioxygen bound to copper during enzyme catalysis.";
 RL Science 286:1724-1728(1999).
 CC -1- CATALYTIC ACTIVITY: 2-phenylethylamine + H(2)O + O(2) =
 CC phenylacetaldehyde + NH(3) + H(2)O(2).
 CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +
 CC H(2)O(2).
 CC -1- COFACTOR: Binds 1 copper ion, 2 calcium ions and 1 topaquinine per
 CC subunit.
 CC -1- PATHWAY: INITIAL STEPS OF 2-PHENYLETHYLAMINE CATABOLISM.
 CC -1- SUBUNIT: HOMODIMER.

CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- PM: Topaquinine (TPQ) is generated by copper-dependent
 CC autoxidation of a specific tyrosyl residue.
 CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D23670; BA004900.1; -;
 DR EMBL: L47571; AAC37012.1; -;
 DR EMBL: AE000235; AAC74468.1; -;
 DR EMBL: D90777; BA14996.1; -;
 DR EMBL: D90776; BA14993.1; -;
 DR EMBL: X97452; CA66104.1; -;
 DR EMBL: X97453; CA66107.1; -;
 DR PDB: 1OAC; 03-MAR-96.
 DR PDB: 1SPU; 12-MAR-97.
 DR PDB: 1QAF; 23-AUG-99.
 DR PDB: 1QAK; 24-AUG-99.
 DR PDB: 1QAL; 24-AUG-99.
 DR PDB: 1DYU; 29-FEB-00.
 DR PDB: 1D6U; 02-FEB-00.
 DR PDB: 1D6Y; 02-FEB-00.
 DR PDB: 1D6Z; 02-FEB-00.
 DR Ecogen; EGI140; CYN.
 DR InterPro; IPR00269; CUNH_oxidase.
 DR Pfam; PF01179; Cu_amine_oxid; 1.
 DR Pfam; PF02727; Cu_amine_oxid; 1.
 DR Pfam; PF02728; Cu_amine_oxid; 1.
 DR PROSITE; PS01164; COPPER_AMINE_OXID_1; 1.
 DR PROSITE; PS01165; COPPER_AMINE_OXID_2; 1.
 DR KX Oxidoreductase; Metal-binding; Copper; Calcium-binding; TPQ;
 KW Periplasmic; Signal; 3D-structure; Complete proteome.
 FT CHAIN 1 30
 FT SIGNAL 1 30
 FT CA_BIND 31 563
 FT CA_BIND 564 564
 FT CA_BIND 565 565
 FT CA_BIND 708 708
 FT CA_BIND 709 709
 FT CA_BIND 603 603
 FT CA_BIND 700 700
 FT CA_BIND 702 702
 FT CA_BIND 709 709
 FT ACT_SITE 413 413
 FT MOD_RES 496 496
 FT METAL 554 554
 FT METAL 556 556
 FT METAL 558 558
 FT METAL 719 719
 FT VARIANT 42 42
 FT VARIANT 59 59
 FT VARIANT 33 33
 FT CONFLICT 248 248
 FT CONFLICT 258 258
 FT CONFLICT 276 276
 FT CONFLICT 288 288
 FT CONFLICT 290 290
 FT CONFLICT 456 456
 FT CONFLICT 659 659
 SO SEQUENCE 757 AA; 84378 MW; 65600BCED35243DB CRC64;
 Query Match 9.7%; Score 397.5; DB 1; Length 757;
 Best Local Similarity 27.5%; Pred. No. 4e-23;
 Matches 142; Conservative 74; Mismatches 211; Indels 89; Gaps 20;
 Oy 265 FYGGRTYDLSLAQLAAGEAGLVNVLIPDNGTGSMS-----LKS 304
 Db 259 YFDGR--DGLKQ-----DARLKVISTLDVGDGYMYNHPINLVAVVDLEQRKIVKIEEG 311


```

QY 305 PVPKPG-----AP---PLOGY-PQGRFVSQGRVASSLWTFSGLGAFSGPR 348
DB 312 PVPVPMPTARPFQDRDRAVAVPMQIEEGKNYITGDMIHRRNDFHLSNMSRVGPM 371
QY 349 IFVVRFG-----BRLVYELQELALAYGNSPAAATRRVYDGS-FQMGKTTPLTGLVD 403
DB 372 ISVTYINDNCKTKKVMYESSLGNTVYGPDPIDGWYFKALDSDGYMGTLSPDIANGKD 431
QY 404 CPYLATYVDWHFLESQAPRTINDAFCEPQONOGPLRLRHSDLSYFYG-----LAET 458
DB 432 AFSNAVLNETIADYGVCPMEIRALVFE-----KAGPEYKQEMQGPVSTERR 483
QY 459 VLVYRSKSTLANTDYWDVYFHPSGALEIRFYATGY-----ISSAFLEGATGK---YGN 509
DB 484 ELVYRMISTGVNDYIFDMIFHENGITIGDAGATGEAVKGVAKTWHDETAADKDTRYGT 543
QY 510 QVSEHTLGTVHTSAHFKVDLVAGLENNVMAEDMVFPAAVWSPRHQRLQDQYTRKL 569
DB 544 LIHNTVGTTHOITVYERLDDVDGENNSLVAMPVVKPMT---AGPRTSTMQVNOYNI 600
QY 570 EMEQOAFVLSGATPRTLYLASNHNKMGHPRGYRIQMLSFAG--EPLPONSSMA----- 622
DB 601 GNEQDAQKDFGTIR-LISNPKENMGNPVS-QIIRYAGTHVAAQAPAPDEMI 657
QY 623 -RQFSERVOLAVTQKKEEPPSSSVF-NONDPAAVTVDFSDPINNETIAGKDLVAVTA 680
DB 658 YHSLFMDKQLMWYRHNHGRFPEGKYPNRSTHDTGLGQYSK--DNESLNTDAVVMWT 715
QY 681 GFLLIHPAEDIPNTVYVNGVGFELRYPNFEDPS 716
DB 716 GTTHVARAEEMPTMP--EMVHTLTKPMNFDEPT 749

RESULT 11
AMOH_ARTGO STANDARD: PRT: 683 AA.
AC 059118:
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Histamine oxidase (EC 1.4.3.6) (Copper amine oxidase).
OS Archaeobacter globiformis.
OC Bacteria: Actinobacteria: Actinobacteria (class): Actinobacteridae;
OC Actinomycetales; Micrococccineae; Micrococaceae; Arthrobacter.
OX NCBI:taxid=165;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-IFO 12137 / ATCC 8010;
RX MEDLINE-95181469; PubMed-7876243;
RA Choi Y.-H., Matsuzaki R., Fukui T., Shimizu E., Yoritani T., Sato H.,
RA Ozaki Y., Tanizawa K.;
RT Copper/topa quinone-containing histamine oxidase from Arthrobacter
RT globiformis. Molecular cloning and sequencing, overproduction of
RT precursor enzyme, and generation of topa quinone cofactor.;
RL J. Biol. Chem. 270:4712-4720(1995).
CC -I- FUNCTION: OXIDIZE HISTAMINE. OTHER AMINES INCLUDING
CC PHEETHYLAMINE, TYRAMINE, TRYPTAMINE, PUTRESCINE, AND BENZYLAMINE
CC ALSO SERVE AS SUBSTRATE.
CC -I- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
CC H(2)O(2).
CC -I- COFACTOR: Binds 1 copper ion and 1 topaquinone per subunit.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- INDUCTION: BY HISTAMINE.
CC -I- PTM: Topaquinone (TPO) is generated by copper-dependent
CC autooxidation of a specific tyrosyl residue.
CC -I- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entries requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL: D8508; BA07517.1; -.
DR HSSP: P46881; 1A4.
DR InterPro: IPR000269; CuNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
DR Pfam: PF02727; Cu_amine_oxid2; 1.
DR Pfam: PF02728; Cu_amine_oxid3; 1.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; FALSE_NEG.
KW Oxidoreductase; Copper; TPO; Metal-binding.
FT INIT_MET 0
FT MOD_RES 401 401 TOPAQUINONE.
FT METAL 450 450 COPPER (POTENTIAL).
FT METAL 452 452 COPPER (POTENTIAL).
FT METAL 611 611 COPPER (POTENTIAL).
SQ SEQUENCE 683 AA; 74979 MW; 233C46C017201AA0 CRC64;

Query Match 9.2%; Score 377; DB 1; Length 683;
Best Local Similarity 24.1%; Pred. No. 1,4e-21;
Matches 182; Conservative 107; Mismatches 343; Indels 124; Gaps 28;

QY 39 PRCPSVPSAQPT--HPGSQLFADLSREELTAVMRFLTORLPGLVDAOARPSDNCV 96
DB 6 PSTPLVQDDPVPATLVHAAQHPLEQSAEIHERRILAE--AGLVG----- 51
QY 97 FSVELQAPRAALAHIDRSPP-----PARALIVFFGRPO-----PVSE 140
DB 52 -----ESTRPATVGLIEPPTTRQGVGTGAARLVRLMWDMAOASRLDVLSTATG 102
QY 141 LVVGPLPHPSYMRDVTYERHGGPLVYHRPVLFOEYLDIDQIMFRELPOASGLHHCCE 200
DB 103 LVVD-----RRELNEAG-----QLVLEEFQIIEIDIL--SDPPQNAL----- 142
QY 201 YKHGRNLVTMTAP-----RGLSGDRATWFGLYNISGAFGLHNVGLLELVNKA 253
DB 143 -TARGLTPACVRAVPASAGVFEYGENEGKRL-----RGLGFRODHP-----ADHPW 188
QY 254 LDPARTIOKVFVQGRYDLSLAQLEQFEGVLNVVYLIP--NCTGSGMSLSKSPVPPAP 312
DB 189 AHFIDGLVAVDVENRRVNL-----IDGPPV--PEVNGYITDPAIRGELRTDLLP 239
QY 313 PLOGYQGRFVSQGRVASSLWTFSGLGAFSGPRIFDVR--FOGER--LYVEISLOEA 368
DB 240 IEIMOPEGPSFTLEGHLSAGWDLRVGFPAKGLVYLQGHSHKGRRRVIRASISEM 299
QY 369 LAIYGNPSPAAATRRVYDGS-FQMGKTTPLTRGVDCPYLATYVDWHFLESQAPRTIR 427
DB 300 VVEYGPSPYRSMQNYFDSGEYLVGRDANSLRGCDLGDITVMSPVVADDFGNPTIEN 359
QY 428 ACFVPEONOGPLRLRHSDLSYFPGSLATVYVNSMSLWLVYDWDVDFHPSGAIET 487
DB 360 GTIHEEDAGILMK--HTDEMAQSDYERRRRLVSEFTVGVYDGFYUWYLVLDGITEF 417
QY 488 RFYATGYISSAFLEGATGKGNQVSEHTLGTVHTSAHFKVDLVAGLENNVMAEDVFEV 547
DB 418 EAKATGIVFTALPDMQVAYASIEAPGLGAPHYOHLESARLDMIDGADRVELDLVRL 477
QY 548 PAAVWSPRHQRLQDQYTRKLEMEQOAFVLSGATPRTLYLASNHS-NKMGHPRGYRIQ 606
DB 478 PKG-PGNPHG--NAFTQKRTLLRESEAVVDADGAGRGVWHISNPDSLNLHGRVGYTL 534
QY 607 MTSFAGEP---LPONSSMARGFSEMYRYOLAVTQKKEEPPSSSVFNONDPAAVTVDFSD 663
DB 535 P--EENPPLAADSSISRAAFARHHLVTRHAAEEELYAADDVFNHGGAVLP-AYV 590
QY 664 INNETIAGDLVAVNAGFLHPIHADIPTVYVNGVGFELRYPNFEDPSFYGASDI 723
DB 591 AQRDIDGQDLVYVHSHFGLHFRPREDW--IMPVDTGFTLKHPPFFDENPPLANVSSA 648
QY 724 YFGDDAGACEVNPACLPQAAACAPDLPAISHG 759

```

Db 649 -----AGHGTG-----SEREHADGGTAVGHSG 672

RESULT 12

AMOI_ARTS1 STANDARD; PRT; 648 AA.

AC 007121;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Copper amine oxidase precursor (EC 1.4.3.6) (MAOXI).
GN MAOI.

OS Arthrobacter sp. (strain P1).
OC Bacteria; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococccineae; Micrococcaceae; Arthrobacter.
OX NCBI_TaxID=47915;

RE MEDLINE=93374858; PubMed=8366046;
RA Zhang X., Fuller J.H., McIntire W.S.;

RT Cloning, sequencing, expression, and regulation of the structural
RT gene for the copper/topa quinone-containing methylamine oxidase from
RT Arthrobacter strain P1, a Gram-positive facultative methylotroph.;

RL J. Bacteriol. 175:5617-5627(1993).

CC -1- FUNCTION: THE EXACT FUNCTION OF MAOXI IS NOT KNOWN.

CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +

CC H(2)O(2).

CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.

CC -1- PTM: Topaquinine (TPO) is generated by copper-dependent

CC autooxidation of a specific tyrosyl residue (by similarity).

CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: L12983; AAA22076.1; -

DR HSSP: P46881; JAV4.

DR InterPro: IPR000269; Cunn_oxidase.

DR Pfam: PF01179; Cu_amine_oxid. 1.

DR Pfam: PF02727; Cu_amine_oxid2. 1.

DR Pfam: PF02728; Cu_amine_oxid3. 1.

DR PROSITE: PS01164; COPPER_AMINE_OXID. 1.

DR PROSITE: PS01165; COPPER_AMINE_OXID. 2.

KW Oxidoreductase; Copper; TPO; Metal-binding.

FT PROPEP 1 9

FT CHAIN 10 648 COPPER AMINE OXIDASE.

FT MOD_RES 385 385 TOPAQUINONE (BY SIMILARITY).

FT METAL 436 436 COPPER (POTENTIAL).

FT METAL 438 438 COPPER (POTENTIAL).

FT METAL 595 595 COPPER (POTENTIAL).

SO SEQUENCE 648 AA; 72760 MW; B2F9E267492253B5 CRC64;

Query Match 8.9%; Score 363; DB 1; Length 648;

Best local Similarity 24.0%; Pred. No. 1,6e-20;

Matches 177; Conservative 96; Mismatches 301; Indels 164; Gaps 31;

55 GOSQFLADLSREELTAVMRELTORLGLVDAAQARSDNCVSEVELQPKAAALALHD 114

12 GVSHPLDPLSRVEIARAVALIKE--GPAASFR-----FISVELREPSKDL----- 57

115 RGSPPAREALAYFFRGOPNPVSELVGPPLPSPTRDVTYRHHGGLPYPHRRPVLEQ 174

58 RAGVAVAREDAVLYV--DRQARSFEAVVD-----LEAGTVSMKLLLENIPPEFLD 108

175 EYLDIDQMIFNRELPOASGLLHCCCFYKHGRNLVTMTAPRGLOSQDRA---TTFGLY 231

109 EFACCEAC--KQDF-----VIALAKRGLTNDLVCPEPMSVGYF 148

QY 232 NISGAFELHHVGLLELVNHNKALD-PARWTIQK--VEYQGRYDSLALQAEAGLVNV 288

Db 149 GEDNEGRRL-MRALVYRDEADSPYAHPIENIVY-----DLNAG--KV 191

QY 289 VLIPDNGTGSWSLSKSPVPPGAPPLQFY-----PGGPRSVQGSRYASS 333

Db 192 VRLDDQA-----IPPSARGNYLPRKYGEARTDLKPLNITQPEGASFTGNHTWA 244

QY 334 LMTSPFLGAFSSGRIDYRQGE-----RLVYEISLQALAIYGNSPA-AMTRYVDG 388

Db 245 DWSFRVGTTPREGVLVLQLFKFKDGVDRPVYINRSLSEMPVYDGTAPVQAKKNAFSGE 304

QY 389 FGMRKTTPLRTGVDCYPLATYVDHMFLESOAKTIRDAFCVFEQNGLLRRHSDLY 448

Db 305 YNIGMANSLTLCGDCGLGEIYFPGHVSVDNGNWTIENALCMHEDDST-----LM 356

QY 449 SH--YFGGLAET---VLVRSNSTLLNYDVTVDYVHPGSAIDIRYANGYISSA-FLF 501

Db 357 KHDFREGTAETRRSRKLVISFIATVANYEYAFVHFLDGSIFLVATGILSTAGQLP 416

QY 502 GATKRYGNQVSEHTL-GVTHSHAFKVDLDVAGLENNVVAEDVFPVMAVPMSPEQL- 559

Db 417 GKRPYQOSLNNDGLVAPRIHOMFNVRWDFELDGVKNAVYEDMEY-----PEHNPT 468

QY 560 -QRLQVTRKLLMEEOAAFLVGSATPRYLYLASNHS-NKMGHPGXYRIQMLSFAGEPLPQ 617

Db 469 GTAFMADRLLETQOKAIRKNEAKHFKTANIESKLVNVEPARYRL-----IPT 519

QY 618 NSSMARGSWERYOLAVYQREKEPSSSVFNQDMPAPYVD-----FSDFINNET-- 668

Db 520 NG-----IOLAA--RDDAVYSKRAOPARNNLWTVADRERRPAAAGEYPMQATGAD 567

QY 669 -----IAGKDLVAVNTAGFLIPIAIEDIPNTVYVNGNVGFLRYNFDDEPS 716

Db 568 DGLHWTOKDRNIVDTDLVWYTFGMHVRLEDP--VMPRONIGFLEPHGFNPNT 625

QY 717 FYSADSIYFRGDODAGAC 734

Db 626 LNLPTSTSTQGTGADRC 643

RESULT 13

AMOI_ARTS1 STANDARD; PRT; 648 AA.

ID AMOI_ARTS1

AC 007123;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Copper methylamine oxidase precursor (EC 1.4.3.6) (MAOXII).

GN MAOI.

OS Arthrobacter sp. (strain P1).

OC Bacteria; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Micrococccineae; Micrococcaceae; Arthrobacter.

OX NCBI_TaxID=47915;

RE MEDLINE=93374858; PubMed=8366046;

RA Zhang X., Fuller J.H., McIntire W.S.;

RT Cloning, sequencing, expression, and regulation of the structural

RT gene for the copper/topa quinone-containing methylamine oxidase from

RT Arthrobacter strain P1, a Gram-positive facultative methylotroph.;

RL J. Bacteriol. 175:5617-5627(1993).

CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +

CC H(2)O(2).

CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.

CC -1- SUBUNIT: HOMODIMER.

CC -1- INDUCTION: BY METHYLAMINE.

CC -1- PTM: Topaquinine (TPO) is generated by copper-dependent

CC autooxidation of a specific tyrosyl residue (by similarity).

CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: L12990; AAA22074.1; -
 DR HSSP: P46881; IAV4.
 DR InterPro: IPR000269; CuNH_oxidase.
 DR Pfam: PF01179; Cu_amine_oxid_1.
 DR Pfam: PF02727; Cu_amine_oxid2; 1.
 DR Pfam: PF02728; Cu_amine_oxid3; 1.
 DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
 DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
 KW Oxidoreductase; Copper; TPQ; Metal-binding.
 FT PROPEP 1 9
 FT CHAIN 10 648 COPPER METHYLAMINE OXIDASE.
 FT MOD_RES 385 385 TOPAQUINONE (BY SIMILARITY).
 FT METAL 436 436 COPPER (POTENTIAL).
 FT METAL 438 438 COPPER (POTENTIAL).
 FT METAL 595 595 COPPER (POTENTIAL).
 SQ SEQUENCE 648 AA; 72805 MW; B2FB2767492253B5 CRC64;

Query Match

Best Local Similarity 24.0%; Score 363; DB 1; Length 648;
 Matches 177; Conservative 96; Mismatches 301; Indels 164; Gaps 31;

QY 55 GOSQLFADLSREELTVMRELTORLPGVDAQAARSDNCVSELTOLPPKAAALALHD 114
 DB 12 GVSHPLDPLSRVETARVALKE--GPAAESR-----FISVELRESKDL----- 57
 QY 115 RGSPPAREALAIFFGROPQPNVSELVGPLRPSYMDVYERRHGRLPYRRPVLFO 174
 DB 58 RAGVAAREADAVLV--DRAQARSFEAYVD-----LEAGTVDSWKLLENIOCPMLD 108
 QY 175 EYLDIQMIFNRELPOASGLHHCFCFKHGRNLVMTTPRGLQSDRA---TWELLY 231
 DB 109 EPECEBAC--RKDPE-----VIAALKRGITNDLVCFEPWSGYF 148
 QY 232 NISGAEFLHVGLELVNHALD-PARKTQK--VFYQGRYDSLQLEAGEAGLVN 288
 DB 149 GEDENERRL--MRALFVVRREADSPYAHPIENPIFY-----DLNAG--KV 191
 QY 289 VLPDNGTGSWSLKSFPVPGAPPLQFY-----PGPRFSVQSGRVASS 333
 DB 192 VRLEDOA-----IPVPSARGNYLPKYVGEARTDLKPLINTQPEASFTVGNHTWA 244
 QY 334 LMFESGLAFSGPRIFDVFEQGE---RLVYELSLQELALAYGNSPA-AMTTRYVDG 388
 DB 245 DMSFRVGFPRDELVLHQLKFKDOGVDRPYNRSLSEMYVPGDTAPVQAKNAFSGE 304
 QY 389 FGAGKYTPPLTRGVDCPYLATYVDWHEFLLESQAPKTRDAFCVFEQNGQLPLRRHSDLY 448
 DB 305 YNIGNMANSLTLCGDGCEIKFYDGHSDVSHGNPMTEINLICHNEEDDSI-----LW 356
 QY 449 SH--YFGGLAET-----VLVRSMSLTLYNDYVMDYFHSGLAIEIRYATGYISSA-FLF 501
 DB 357 KHFDFREGTAETRSRLVLSFATVANYEAFAVWHLFLDGSIEFLVAKATGILSTAQLP 416
 QY 502 GAGGKGNQVSEHTL-GTYVTHSHAFKVDVADVAGLEWMAEDVFPVPAVWSPHQQL- 559
 DB 417 GEKNPFGQSLNDGLVPIQHMFNVMDELDGVKNAYVEVDEY-----PEHNPT 468
 QY 560 -ORLOYTRKLLMEEQAAFLVGSATPRYLTLASNHS--NKMGHPRGYRIQMLSFAGEPLPQ 617
 DB 469 GTAFMAVDRLEETQKAIKRTNEAKHFKIANHESKNLVNEPVAIYL-----IPT 519
 QY 618 NSSMARGFSWERYQLAVTORKEEPPSSSVFNQNDPAAPIYD-----FSDFINNET-- 668
 DB 520 NG-----IQLAA--RQDAVYSKRAQFARNMLWTAVDRTERFAAGSEYPMQATGAD 567
 QY 669 -----IAGKLVAVNTAGFLIHPAIEDIPNTVYGVNGVGFLLRYNFFDEPS 716

DB 568 DGLHIWTKDRNIVDIDLVVMTFGMHVVRLEDMP--VAPRONIGFMLEPHGFENONPT 625
 QY 717 FYSADSIYFRGDODAGAC 734
 DB 626 LNPFTSTSTTQTGEADTC 643

RESULT 14
 ID PAOX_ARTGO STANDARD: PRT: 638 AA.
 AC P46881;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Phenylethylamine oxidase precursor (EC 1.4.3.6) (Amine oxidase).
 OS Arthrobacter globiformis.
 OC Bacteria; Actinobacteria; Actinobacteriales (class); Actinobacteridae;
 OC Actinomycetales; Micrococcales; Micrococcaceae; Arthrobacter.
 OX NCBI_TaxID=1665;
 RN (1)
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=IFO 12137 / ATCC 8010.
 RX MEDLINE=94197690; PubMed=8147851;
 RA Tanizawa K., Matsuzaki R., Shimizu E., Yoritani T., Fukui T.;
 RT Cloning and sequencing of phenylethylamine oxidase from Arthrobacter
 RT globiformis and implication of Tyr-382 as the precursor to its
 RT covalently bound quinone cofactor.
 RL Biochem. Biophys. Res. Commun. 199;1096-1102(1994).
 RN (2)
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 9-628.
 RX MEDLINE=98070236; PubMed=9405045;
 RA Wlcek M.C., Doolley D.M., Freeman H.C., Guss J.M., Matsunami H.,
 RA McIntire W.S., Ruggiero C.E., Tanizawa K., Yamaguchi H.;
 RT Crystal structures of the copper-containing amine oxidase from
 RT Arthrobacter globiformis in the holo and apo forms: implications for
 RT the biogenesis of topaquinone.
 RL Biochemistry 36;16116-16133(1997).
 CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +
 CC H(2)O(2).
 CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinone per subunit.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- INDUCTION: BY PHENETHYLAMINE.
 CC -1- PTM: Topaquinone (TPQ) is generated by copper-dependent
 CC autoxidation of a specific tyrosyl residue.
 CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: U03517; AAA18114.1; -
 DR PDB: IAVK; 18-MAR-98.
 DR PDB: IAVL; 18-MAR-98.
 DR PDB: IAV4; 25-MAR-98.
 DR InterPro: IPR000269; CuNH_oxidase.
 DR Pfam: PF01179; Cu_amine_oxid_1.
 DR Pfam: PF02727; Cu_amine_oxid2; 1.
 DR Pfam: PF02728; Cu_amine_oxid3; 1.
 DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
 DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
 KW Oxidoreductase; Copper; TPQ; 3D-structure; Metal-binding.
 FT PROPEP 1 2
 FT CHAIN 3 638 PHENYLETHYLAMINE OXIDASE.
 FT MOD_RES 382 382 TOPAQUINONE.
 FT METAL 431 431 COPPER.
 FT METAL 433 433 COPPER.
 FT METAL 592 592 COPPER.
 FT DISULFID 317 343
 FT MOTAGEN 382 382

Y->F: LOSS OF ACTIVITY.

50 SEQUENCE 638 AA; 70646 MW; 18003968A7A983F2 CRC64;

Query Match 8.3%; Score 340; DB 1; Length 638;

Best Local Similarity 23.7%; Pred. No. 9.7e-10; Matches 157; Conservative 100; Mismatches 274; Indels 132; Gaps 28;

106 KAAALAHLD--RGSPPARALAIVF---GROPQNVSELVVGPLPHSPMDVTER 159
 38 RIATVGLDPAKRGASGEARFRVFIHDVSGARQ-----EYTVSV 79

160 HGGFL-----PYHRPVLFEYLDIDOMFN-----RELPOAS 192

80 TNGTVISAVELDTATGELPLTEEFEVEEQLLATDERMLKALARNIDVSKVRAPLSA 139

193 GLTHRCCEYKHKRNLYMTAPRGL---QSGRATWGLYXINISGAFGLHHVGLLE 248

140 GVEFYA---ERGRRL-----RGLAVODPPEDSAN---AHPDGLVAIVDVVSKVET 187

249 VNRKALDPAWTKIVKYQGRYDLSAQLAQAFAGLVNVVLIIDNCTGSSMLKSPVP 308

188 --RVIDTGVPRVA--EHGNYTDP-----ELTGLPRT 215

309 GRAPPLQYPOGPRFSVO--GSRVASSLMTFSFGIGASGPRIFDVRO--GERL---VYEI 363

216 TOKRISTQPEGPSFTYTGNNHMEKMSLDVGFVREGVVLHNIAFRDGRLRPILINRA 275

364 SLOALAIYGGNSPAAATTRYVD--GGFGMGKYYTPTLRGVGD---PYLATYVDMHLE 418

276 SIAMVYPRGDPSPRISMONTFDGEYLVGOYANSLEIGCDLGDITLSVYISDAF--- 332

419 SQAPKTRDAFCVEQNOGLPLRRHSDLSHYFGIAETVLVYRSMSTLNYDVMDTV 478

333 -GNREIRNGICMEHEDMGLAK--HSDLMGINVTRNRNRVMSFFITIGNYDGYFWY 389

479 FHFGAIEIRYATGYI--SSAFLEGATGKYNOVSEHTLGVTHSHAFKVDLDVAGLEN 537

390 LYLDGTLEFKAAGVFTSAFPEGGSNT--SOLAPGIGAFPHOHSARLMDAIDGFTN 448

538 WYNAEDVYFMAVWSPREHOLQVTRKLLMEEQAAFLVGSATPYKLVIASNHS--NK 596

449 RVEEDVYRQMGNGRGNFNR---KRYVLTRESEVNRADARTGTWITSNPESNKR 505

597 WGRGRGRI---OMLSFAGEPLPONSSMARGESWERYOLAVTORKKEEPPSSSVF--NOND 652

506 LNEPVGKYLHANOPLTLPD--GSSIAARAFAKTDLMVTRADDERVPTGDVFNHGS 562

653 PMAPTVDSPFINNETIAGKDLVAVMTAGFLHHAEDIPVTVVGVNGVGFLLAPYNED 712

563 GGAGLPSY--IAQDRDIDGODIVVMTFTGLTHPRVEDMP--IMPVDTVGKLRPEGFED 618

713 EDP 715

619 RSP 621

RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RX MEDLINE=9639683; PubMed=8805580.
 RA Kumar V., Dooley D.M., Freeman H.C., Guss J.M., Harvey I.,
 RA Mcguil M.A., Wilce M.C., Zubak V.M.;
 RT "Crystal structure of a eukaryotic (pea seedling) copper-containing
 RT amine oxidase at 2.2-A resolution.",
 RL Structure 4:943-955(1996).
 CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
 CC H(2)O(2).
 CC -1- COFACTOR: Binds 1 copper ion, 1 manganese ion and 1 topaquinine
 CC per subunit.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- PPM: Topaquinine (TPQ) is generated by copper-dependent
 CC autooxidation of a specific tyrosyl residue (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L39931; AAA62490.1; -
 DR PDB: 1KST; 24-DEC-97.
 DR InterPro: IPR000269; CuNH_oxidase.
 DR Pfam: PF01179; Cu_amine_oxid_1.
 DR Pfam: PF02727; Cu_amine_oxid2; 1.
 DR Pfam: PF02728; Cu_amine_oxid3; 1.
 DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
 DR PROSITE: PS01165; COPPER_AMINE_OXID_2; FALSE_NEG.
 KW Oxidoreductase; Metal-binding; Copper; Manganese; TPQ; Glycoprotein;
 FT Signal; 3D-structure.
 FT SIGNAL 1
 FT CHAIN 26 674
 FT METAL 476 476 AMINE OXIDASE [COPPER-CONTAINING].
 FT METAL 477 477 MANGANESE (VIA CARBONYL OXYGEN).
 FT METAL 478 478 MANGANESE (VIA CARBONYL OXYGEN).
 FT METAL 617 617 MANGANESE.
 FT METAL 618 618 MANGANESE (VIA CARBONYL OXYGEN).
 FT MOD_RES 412 412 TOPAQUINONE.
 FT METAL 467 467 COPPER.
 FT METAL 469 469 COPPER.
 FT METAL 628 628 COPPER.
 FT DISULFID 162 183
 FT DISULFID 344 370
 FT CARBOHYD 156 156 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CARBOHYD 389 389 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CARBOHYD 583 583 N-LINKED (GLCNAC. . .) (PROBABLE).
 SQ SEQUENCE 674 AA; 76358 MW; 30735390071DD18E CRC64;

Query Match 7.8%; Score 320; DB 1; Length 674;
 Best Local Similarity 25.8%; Pred. No. 3.8e-17;
 Matches 120; Conservative 66; Mismatches 189; Indels 90; Gaps 20;

301 SLKSPVPRGAPP--LQFYQGPFRFSVQSGRVASSLMTFSFGLCARFG-----PRIDV- 352
 233 SKOSP-PFGKQSHSLTSHOQPGFOJNGHSVANKRFHIGEDVARGIVISLASIVLE 291
 353 RFOGERLVEISLOEALAIYGGNSPAAATTRYVCG-FGAGKYYTPTLRGVDCDPLATYV 411
 292 KHKRRVLYKGYISELVVPODPREEYFTTFPDSGFGGLSTVSLIPNDCPHAOFT 351
 412 DMHFLSEAPKTRDAFCVEQNOGLPLRRHSDLSHYFGGL-----AETVLV 462
 352 DTYVHSANGTPILLKNKICVFEQ-----YGINMRHTEINGIPNSIESRREVALIV 403
 463 RSMSTLNYDVMDTVYHFGAIEIRYATGYISSAFLFGATGK-----YGNQVSE 513
 404 RTIVTVGNVNDVIDWPKASGISKPSIALSGILE--IKGTNKKHDEIKEDLHGKLVSA 460

RESULT 15
 AMO_PEA STANDARD; PRT; 674 AA.
 AC 043077;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Amino oxidase [copper-containing] precursor (EC 1.4.3.6).
 OS Pisum sativum (Garden pea).
 OC Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
 ON NCBI_TaxID=3888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tipping A.J., McPherson M.J.;
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]

Tue May 20 13:18:55 2003

us-10-081-408-2.rsp

Page 15

[illegible]

Search completed: May 20, 2003, 11:34:58
Job time : 22.01 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 10:12:24 ; Search time 52.6446 Seconds

(without alignments)
2986.328 Million cell updates/sec

Title: US-10-081-408-2

Perfect score: 4095

Sequence: 1 MNOKTILVLLAVITFAL.....QAACAPDLPAFSGFSHN 763

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organelle:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp.todent:*
- 12: sp.virus:*
- 13: sp.vertibrate:*
- 14: sp.unclassified:*
- 15: sp.virus:*
- 16: sp.bacteriap:*
- 17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3517	85.9	763	6	Q9TKK6	Q9TKK6 Bos taurus
2	3478	84.9	765	11	Q9R055	Q9R055 mus musculus
3	1404.5	34.3	751	11	Q8VC36	Q8VC36 mus musculus
4	1032	25.2	447	11	Q8R229	Q8R229 mus musculus
5	649	15.8	787	3	Q96X16	Q96X16 pichia past
6	615.5	15.0	271	11	Q920K6	Q920K6 cavia porce
7	472.5	11.5	204	11	Q9CWL6	Q9CWL6 mus musculus
8	385	9.4	741	10	Q9C6V7	Q9C6V7 arabidopsis
9	381	9.3	654	16	Q889X2	Q889X2 rhizobium
10	372.5	9.1	668	10	Q48552	Q48552 arabidopsis
11	368.5	9.0	660	17	Q97XW1	Q97XW1 sulfolobus
12	358	8.7	64	4	Q9UEU7	Q9UEU7 homo sapien
13	357.5	8.7	756	10	Q9STI2	Q9STI2 arabidopsis
14	357	8.7	194	4	Q9HBM7	Q9HBM7 homo sapien
15	353.5	8.6	650	10	Q23349	Q23349 arabidopsis
16	353	8.6	649	10	Q8W1C1	Q8W1C1 brassica ju

17	343.5	8.4	735	10	Q9SW88	Q9SW88 canavalia l
18	338	8.3	1794	10	Q9S168	Q9S168 arabidopsis
19	335.5	8.2	422	10	Q9C6M1	Q9C6M1 arabidopsis
20	328	8.0	674	10	Q9SXW5	Q9SXW5 pisum sativ
21	327	8.0	687	10	Q9M2B9	Q9M2B9 arabidopsis
22	320.5	7.8	670	10	Q65749	Q65749 cicer ariet
23	319	7.8	420	10	Q9C6V9	Q9C6V9 arabidopsis
24	318	7.8	701	10	Q9XHP4	Q9XHP4 glycine max
25	310	7.6	677	10	Q9SW90	Q9SW90 euphorbia c
26	309	7.5	667	10	Q9LD03	Q9LD03 lens culina
27	307.5	7.5	638	3	Q96X06	Q96X06 aspergillus
28	302.5	7.4	759	10	Q9SLB1	Q9SLB1 arabidopsis
29	300	7.3	660	16	Q8YRL5	Q8YRL5 arabidopsis
30	279.5	6.8	712	3	Q9P7F2	Q9P7F2 schizosacch
31	257.5	6.3	118	11	Q9DWM4	Q9DWM4 mus musculu
32	247	6.0	794	3	Q42890	Q42890 schizosacch
33	246.5	6.0	300	10	Q9STI3	Q9STI3 arabidopsis
34	200	4.9	103	11	Q63973	Q63973 rattus norv
35	179.5	4.4	102	6	Q29317	Q29317 sus scrofa
36	126	3.1	862	4	Q9WT23	Q9WT23 homo sapien
37	125.5	3.1	460	10	Q9STI4	Q9STI4 arabidopsis
38	125	3.1	757	16	Q9K6G6	Q9K6G6 bacillus ha
39	115	2.8	2675	2	Q9FB23	Q9FB23 streptomyce
40	114.5	2.8	949	5	Q952L5	Q952L5 caenorhabd
41	113.5	2.8	1275	6	Q62658	Q62658 canis fami
42	113	2.8	43	11	Q99NA5	Q99NA5 cavia porce
43	113	2.8	788	5	Q8STB6	Q8STB6 desophila
44	112.5	2.7	605	3	Q96VA3	Q96VA3 aspergillus
45	112	2.7	470	17	Q8ZYX8	Q8ZYX8 pyrobaculum

ALIGNMENTS

RESULT 1
ID Q9TKK6 PRELIMINARY: PRT: 763 AA.

AC 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-MAR-2002 (TREMREL. 20, Last annotation update)
DE Semicarbazide-sensitive amine oxidase (EC 1.4.3.6).
GN SSO.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Iwabuchi H., Matsumura K., Mure M., Kuroda S., Tanizawa K.;
RT "Molecular cloning of semicarbazide-sensitive amine oxidase gene from
RT Bovine aorta.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB019242; BAA8896.1; -;
DR InterPro: IPR000269; Cunn_oxidase.
DR Pfam: PF01179; Cu_amine_oxid. 1.
DR Pfam: PF02727; Cu_amine_oxid2; 1.
DR Pfam: PF02728; Cu_amine_oxid3; 1.
DR PRINTS: PR00766; CUDAOXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID.1; UNKNOWN.1.
DR PROSITE: PS01165; COPPER_AMINE_OXID.2; 1.
KW oxidoreductase.
SQ
SEQUENCE 763 AA: 84500 MW: DBE1ED9A89E71E90 CRC64:

Query Match 85.9%; Score 3517; DB 6; Length 763;
Best local similarity 85.2%; Pred. No. 3.2e-280;
Matches 650; Conservative 44; Mismatches 69; Indels 0; Gaps 0;

QY 1 MNOKTILVLLAVITFALVCLVGRGGGEPQLPHCPSPVSPAQPMTHPGOSQLF 60
DB 1 MNOKTILVLLAVITFALVCLVGRGGGEPQLPHCPSPVSPAQPMTHPGOSQLF 60

```
OY 61 ADLSREELTAVMRELTORLPGVLDAAQARPNDNCVSEVLEQLPPKAAALAHLDRGSP 120
DB 61 ADLSREELTAVMRELTORLPGVLDAAQARPNDNCVSEVLEQLPPKAAALAHLDRGSP 120
OY 121 AREALATVFCFQOPVYSELVGLPLPHPSYMRDVTYERHGGPLPYHRRPVLPQEVYDID 180
DB 121 AREALATVFCFQOPVYSELVGLPLPHPSYMRDVTYERHGGPLPYHRRPVLPQEVYDID 180
OY 121 AREALATVFCFQOPVYSELVGLPLPHPSYMRDVTYERHGGPLPYHRRPVLPQEVYDID 180
DB 121 AREALATVFCFQOPVYSELVGLPLPHPSYMRDVTYERHGGPLPYHRRPVLPQEVYDID 180
OY 181 QMIFNRELPOASGLLHHCCKYKRGRLVMTTAPRGLQSDRATWGLYNNISGAFEL 240
DB 181 QMIFNRELPOASGLLHHCCKYKRGRLVMTTAPRGLQSDRATWGLYNNISGAFEL 240
OY 241 HHVGLLELVNHNKALDPARMTIOKRYGYRYSLSLAOLEAOEALVNVVLPDNGTGGSW 300
DB 241 HHVGLLELVNHNKALDPARMTIOKRYGYRYSLSLAOLEAOEALVNVVLPDNGTGGSW 300
OY 301 SLKSPVPPGAPPLQFYPQGRFVSQGSRYVASSLMTFSFGAGSGRIPDRVROGERLY 360
DB 301 SLKSPVPPGAPPLQFYPQGRFVSQGSRYVASSLMTFSFGAGSGRIPDRVROGERLY 360
OY 301 SLKSPVPPGAPPLQFYPQGRFVSQGSRYVASSLMTFSFGAGSGRIPDRVROGERLY 360
DB 301 SLKSPVPPGAPPLQFYPQGRFVSQGSRYVASSLMTFSFGAGSGRIPDRVROGERLY 360
OY 361 YEISLOALAIYGNNSPAAATTRYVDGFGNGKATTPTRGVDCPYLATYVDHMFLESO 420
DB 361 YEISLOALAIYGNNSPAAATTRYVDGFGNGKATTPTRGVDCPYLATYVDHMFLESO 420
OY 421 APKTRDAFCVFEONOGPLRRHNSDLXSHYFGGLAEVLYVRSMTLLANTDYMDVYFH 480
DB 421 APKTRDAFCVFEONOGPLRRHNSDLXSHYFGGLAEVLYVRSMTLLANTDYMDVYFH 480
OY 481 PSGLAIEIRFATGYISSAFLEGATGKYGNVSEHTLCTVHTSHAHFVYDLVAGLENNVW 540
DB 481 PSGLAIEIRFATGYISSAFLEGATGKYGNVSEHTLCTVHTSHAHFVYDLVAGLENNVW 540
OY 541 AEDVAFVPMAPMSPEHOLQRLQVTRKLEMEBOAFLVGSATPRYLALSNHNSKNGHP 600
DB 541 AEDVAFVPMAPMSPEHOLQRLQVTRKLEMEBOAFLVGSATPRYLALSNHNSKNGHP 600
OY 601 RGYRIQMLSFAGEPLPONSSMARGFSWERYOLAVTORKEEPPSSSVFNQNDPAPYVDF 660
DB 601 RGYRIQMLSFAGEPLPONSSMARGFSWERYOLAVTORKEEPPSSSVFNQNDPAPYVDF 660
OY 661 SDFINNETIACKDLVAVWTAGFLIIPHAEDIPNVTYVNGVGFELRPVNFEDDEPSTYSA 720
DB 661 SDFINNETIACKDLVAVWTAGFLIIPHAEDIPNVTYVNGVGFELRPVNFEDDEPSTYSA 720
OY 721 DSIFRGDODGACEVNPPLACTPOAAACAPDLPAFSGHGFESH 763
DB 721 DSIFRGDODGACEVNPPLACTPOAAACAPDLPAFSGHGFESH 763

RESULT 2
OY 09R055 PRELIMINARY: PRT: 765 AA.
AC 09R055;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAY-2002 (Tremblrel. 20, Last annotation update)
DE Copper amine oxidase (EC 1.4.3.6).
CN AOC3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99194802; PubMed=10092636;
RA Moltes M., Feve B., Paltrout J.;
RT "Molecular cloning of a major mRNA species in murine 3T3 adipocyte
RT lineage: differentiation-dependent expression, regulation, and
RT identification as semicarbazide-sensitive amine oxidase.";
RL J. Biol. Chem. 274:9513-9523(1999).
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
CC H(2)O(2).
CC -1- COFACTOR: COPPER AND TOPAQUINONE (BY SIMILARITY).
```

```
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
DR EMBL: AF115411; MAD09199.1; -.
DR MGD: MG11306797; AOC3.
DR InterPro: IPR000269; CuNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid_1.
DR Pfam: PF02727; Cu_amine_oxid2_1.
DR Pfam: PF02728; Cu_amine_oxid3_1.
DR PRINTS: PR00766; CUDAOXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
DR KEGG: OXIDOREDUCTASE; TPO.
SQ SEQUENCE 765 AA; 84534 MW; 4252854759F316C7 CRC64;
```

```
Query Match 84.9% Score 3478; DB 11; Length 765;
Best Local Similarity 83.3% Pred. No. 5,1e-277;
Matches 635; Conservative 60; Mismatches 67; Indels 0; Gaps 0;
```

```
OY 1 MNOKTILVLLIATVITFALVCVLLVGRGSGSPQLPHCPSPVSPAOPWTHGQSOLF 60
DB 1 MNOKTILVLLIATVITFALVCVLLVGRGSGSPQLPHCPSPVSPAOPWTHGQSOLF 60
OY 61 ADLSREELTAVMRELTORLPGVLDAAQARPNDNCVSEVLEQLPPKAAALAHLDRGSP 120
DB 61 ADLSREELTAVMRELTORLPGVLDAAQARPNDNCVSEVLEQLPPKAAALAHLDRGSP 120
OY 121 AREALATVFCFQOPVYSELVGLPLPHPSYMRDVTYERHGGPLPYHRRPVLPQEVYDID 180
DB 121 AREALATVFCFQOPVYSELVGLPLPHPSYMRDVTYERHGGPLPYHRRPVLPQEVYDID 180
OY 181 QMIFNRELPOASGLLHHCCKYKRGRLVMTTAPRGLQSDRATWGLYNNISGAFEL 240
DB 181 QMIFNRELPOASGLLHHCCKYKRGRLVMTTAPRGLQSDRATWGLYNNISGAFEL 240
OY 241 HHVGLLELVNHNKALDPARMTIOKRYGYRYSLSLAOLEAOEALVNVVLPDNGTGGSW 300
DB 241 HHVGLLELVNHNKALDPARMTIOKRYGYRYSLSLAOLEAOEALVNVVLPDNGTGGSW 300
OY 301 SLKSPVPPGAPPLQFYPQGRFVSQGSRYVASSLMTFSFGAGSGRIPDRVROGERLY 360
DB 301 SLKSPVPPGAPPLQFYPQGRFVSQGSRYVASSLMTFSFGAGSGRIPDRVROGERLY 360
OY 361 YEISLOALAIYGNNSPAAATTRYVDGFGNGKATTPTRGVDCPYLATYVDHMFLESO 420
DB 361 YEISLOALAIYGNNSPAAATTRYVDGFGNGKATTPTRGVDCPYLATYVDHMFLESO 420
OY 421 APKTRDAFCVFEONOGPLRRHNSDLXSHYFGGLAEVLYVRSMTLLANTDYMDVYFH 480
DB 421 APKTRDAFCVFEONOGPLRRHNSDLXSHYFGGLAEVLYVRSMTLLANTDYMDVYFH 480
OY 481 PSGLAIEIRFATGYISSAFLEGATGKYGNVSEHTLCTVHTSHAHFVYDLVAGLENNVW 540
DB 481 PSGLAIEIRFATGYISSAFLEGATGKYGNVSEHTLCTVHTSHAHFVYDLVAGLENNVW 540
OY 541 AEDVAFVPMAPMSPEHOLQRLQVTRKLEMEBOAFLVGSATPRYLALSNHNSKNGHP 600
DB 541 AEDVAFVPMAPMSPEHOLQRLQVTRKLEMEBOAFLVGSATPRYLALSNHNSKNGHP 600
OY 601 RGYRIQMLSFAGEPLPONSSMARGFSWERYOLAVTORKEEPPSSSVFNQNDPAPYVDF 660
DB 601 RGYRIQMLSFAGEPLPONSSMARGFSWERYOLAVTORKEEPPSSSVFNQNDPAPYVDF 660
OY 661 SDFINNETIACKDLVAVWTAGFLIIPHAEDIPNVTYVNGVGFELRPVNFEDDEPSTYSA 720
DB 661 SDFINNETIACKDLVAVWTAGFLIIPHAEDIPNVTYVNGVGFELRPVNFEDDEPSTYSA 720
OY 721 DSIFRGDODGACEVNPPLACTPOAAACAPDLPAFSGHGFESH 762
DB 721 DSIFRGDODGACEVNPPLACTPOAAACAPDLPAFSGHGFESH 762

RESULT 3
OY 08VC36 PRELIMINARY: PRT: 751 AA.
DB 08VC36
```


AC 08VC36:
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Similar to anilotide binding protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=COLON;
 RA Strausberg R.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC021880; AAH21880.1;
 DR InterPro; IPR003439; ABC_transport.
 DR InterPro; IPR000269; CUNH_oxidase.
 DR Pfam; PF01179; Cu_amine_oxid_1.
 DR Pfam; PF02727; Cu_amine_oxid_2; 1.
 DR Pfam; PF02728; Cu_amine_oxid_3; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN.1.
 DR PROSITE; PS01164; COPPER_AMINE_OXID_1; UNKNOWN.1.
 DR PROSITE; PS01165; COPPER_AMINE_OXID_2; UNKNOWN.1.
 SQ SEQUENCE 751 AA; 85449 MW; 178B430795A2B988 CRC64;

Query Match 34.3%; Score 1404.5; DB 11; Length 751;
 Best Local Similarity 39.8%; Pred. No. 1.4e-106;
 Matches 298; Conservative 127; Mismatches 270; Indels 53; Gaps 15;

OY 45 SPSAQPWTHPGOSOLFADLSREELTAVMFLQRLGPGLYDAQAQSPSCNCFVSLELOP 104
 DB 19 TPSAVTTPH-DKARFADISPOEIKAVHSFLMSKELGLSSKNITLANSVFLIEMLP 77
 OY 105 PKAALAHLDHGRSPPARALAIVEFGROPQPNVELVYGPLPHPSMDVVERHGP 164
 DB 78 KKNVTLKFLDEGRKSPVREARAIIFGAGDHNVTEFANGLPRCIYQALS-PRGHIL 136
 OY 165 PYHRRPVLEFOEYLDIDQIFNELPQASGLHHCPCYKHRG-----RNLVTMTADR 216
 DB 137 SWSSPRISTAEY-DLLYHMLNRAITPL-----HOFFLDITGFSFLGCDREFLFTDVAPR 190
 OY 217 GLQSDRAVTFGLYNNISAGFELHNVGELLVNHKALDPAWITQKXVYQADSLAQ 276
 DB 191 GYESQGRSMIVQRYVE--GTFLEPTGLEILYDHSSTDVQDMRVQLNCKFTNSPPE 248
 OY 277 LEAEGAGLVNVVLLPDNGTGSMSLKSPVPG-PAPPL-----QFY----- 317
 DB 249 LAQKAVGEVAVLEEE-----VLEEDPLPGATBOPPLFSSYKPRGEHPTVYAGPIV 302
 OY 318 --PQPRPSVQGRVASSLMTFSFGLGAFSGPRIDVRFQGERLYVEISLOEALAYG 375
 DB 303 VQSPSPRKLLEGNNVLYGDMFSYRLRSSGLQIFNVLEGGERVAYEVSQAVVLYLGGH 362
 OY 376 SPAATTRIVDGFQMGKATYTPLTGRGVDCPYLATYVDHMFLLSQAOKTIRDAFCVEON 435
 DB 363 TPAGMOTKIDYQMGKGLSTHGLAIDCEPETATFIDAHYHDSQGVLYPRALCLFEAP 422
 OY 436 OGLPLRRHSDLYS--HYFGGLAETVLVVRSMSTLLNDYVMDVTFHPSGALIEIRFAT 492
 DB 423 TGVPLRRHSDLYS--HYFGGLAETVLVVRSMSTLLNDYVMDVTFHPSGALIEIRFAT 492
 OY 433 GYISSAFLEGATGKXGNVSEHTLGTVTHSAHFQVDLDVAGLENNVMAEDMVFPMANP 552
 DB 483 GYVAHATFTPEGLRHGTRQLQTHLGNITHLVHVRVDLVAGTAKNSFRTLKLENTINP 542
 OY 553 WSPHOLRLQVTRKLEMEBOAFLVGSATPRVLYLASNSNKGHPRGYRIOMLSFAG 612
 DB 543 WSPHSLVQPTLEQTOYSHHQAAFRFGQTLPKYLLFSSPQKNRKHRSIRLIQHSMAE 602
 OY 613 EPLPONSSMARGFSEMYQOLAVTORKEEPPSSSVFNQNDMPAPYDVSDFI-NNETIAG 671
 DB 603 QVLPFGMWQGERAVVTARVYPLAVTKYRESERYSSSLVYNDQNDPMDPVPVVEEPLRNENIN 662

OY 672 KDLYAMTAGLHIHPHADIPNTVTGNGVGFELRPYNFEDEDPSSYASDIYFGDDA 731
 DB 663 EDLYAMVTVGFHHPHSEDPNTATPGNCVGLIRPFNFEEBPSLASRDYIV-WPDQ 721
 OY 732 GACEVNPPLACLPQAACAAPDLPAFSGH 759
 DB 722 GUNHQ--RWIPENRDCVLS-PPFSYNG 746

RESULT 4

OY 08R229 PRELIMINARY; PRT: 447 AA.

AC 08R229:
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 51.3 kDa protein (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=COLON;
 RA Strausberg R.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC022627; AAH22627.1;
 DR KW Hypothetical protein.
 FT NON-TER
 SQ SEQUENCE 447 AA; 51265 MW; E556EAA2525D1D1 CRC64;

Query Match 25.2%; Score 1032; DB 11; Length 447;
 Best Local Similarity 44.6%; Pred. No. 2.8e-76;
 Matches 199; Conservative 78; Mismatches 161; Indels 8; Gaps 5;

OY 318 PQGPRPSVQGRVASSLMTFSFGLGAFSGPRIDVRFQGERLYVEISLOEALAYG 377
 DB 1 PSGPRKLEGNVLYGDMFSYRLRSSGLQIFNVLEGGERVAYEVSQAVVLYLGGH 60
 OY 378 AAMTRYVDGFGMGKATYTPLTGRGVDCPYLATYVDHMFLLSQAOKTIRDAFCVEON 437
 DB 61 ACGMOTKIDYQMGKGLSTHGLAIDCEPETATFIDAHYHDSQGVLYPRALCLFEAP 120
 OY 438 LPLRRHSDLYS--HYFGGLAETVLVVRSMSTLLNDYVMDVTFHPSGALIEIRFAT 494
 DB 121 VPLRRHSDLYS--HYFGGLAETVLVVRSMSTLLNDYVMDVTFHPSGALIEIRFAT 494
 OY 495 ISSAFLEGATGKXGNVSEHTLGTVTHSAHFQVDLDVAGLENNVMAEDMVFPMANP 554
 DB 181 VHAATFTPEGLRHGTRQLQTHLGNITHLVHVRVDLVAGTAKNSFRTLKLENTINP 540
 OY 555 WSPHOLRLQVTRKLEMEBOAFLVGSATPRVLYLASNSNKGHPRGYRIOMLSFAG 614
 DB 241 PHSLSVQPTLEQTOYSHHQAAFRFGQTLPKYLLFSSPQKNRKHRSIRLIQHSMAE 300
 OY 615 LPONSSMARGFSEMYQOLAVTORKEEPPSSSVFNQNDMPAPYDVSDFI-NNETIAG 673
 DB 301 LPPGMOEERAVVTARVYPLAVTKYRESERYSSSLVYNDQNDPMDPVPVVEEPLRNENIN 660
 OY 674 LYAMTAGLHIHPHADIPNTVTGNGVGFELRPYNFEDEDPSSYASDIYFGDDA 733
 DB 361 EDLYAMVTVGFHHPHSEDPNTATPGNCVGLIRPFNFEEBPSLASRDYIV-WPDQ 721
 OY 734 CEVNPPLACLPQAACAAPDLPAFSGH 759
 DB 420 NHVQ--RWIPENRDCVLS-PPFSYNG 746

RESULT 5
 OY 096X16 PRELIMINARY; PRT: 787 AA.
 AC 096X16:
 DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT	01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT	01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT	01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE	lysyl oxidase.
GN	ADCI.
OS	Pichia pastoris (Yeast).
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC	Saccharomycetales; Saccharomycetaceae; Pichia.
OX	NCBI_TaxId=4922;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=2113196; Pubmed=11237259;
RA	Kucha J.A., Dooley D.M.;
RT	"Cloning, sequence analysis, and characterization of the 'lysyl
RT	oxidase' from Pichia pastoris."
RL	J. Inorg. Biochem. 83:193-204(2001).
DR	EMBL: AF358434; AAK49976.1; "
DR	InterPro: IPR000269; CUNH_oxidase.
DR	Plam; PFO1179; Cu_amine_oxid; 1.
SO	SEQUENCE 787 AA; 89690 MW; D779B8F2BF766745 CRC64;
Query Match	15.8% Score 649; DB 3; Length 787;
Best Local Similarity	27.3%; Pred. No. 2e-44;
Matches 201; Conservative 114; Mismatches 320; Indels 102; Gaps 21.	
QY	45 SPSAOWTHPGOSQLEADLSREELTAVMFYLQRLGLVDAQAQAPSDNCVFSVELQRP 104
Db	55 APTKNTMT-----SLAKEEVOEVLDDLHSTY--NITEVTKADFSNMYLMIETLKP 103
QY	105 PKAALALDLDRGSPRAREALATVFFGQRPQPNVSELVYGLPHHSYMDVTVENHGCL 164
Db	104 NKTLEALITLDEDDGLPPRNARTVYFEGGEDEYFEBELKVP---VSDETIE---PL 156
QY	165 PYHRRPVLFOEYLDIDOMIFNRELPOASGLNHCCFYKRGHNLVTMTAPRGL----- 218
Db	157 SFYNNNGSKSLRPEVGHIDRIKSAKS-----FLKNKNTMTIMRDVLEGLGVPE 208
QY	219 -----QGDRATWEGLYY-----NISGAGFLHNVGLELNVHKAALDPA 257
Db	209 DMGCHSAAPQLHDPATGATVDGTGTCNINTENDAEMLVPTGEFFK-----FDMTGRDVS 261
QY	258 RMTIOKVYQGRKYVYSLAQL-EAQFEAGLVNVVLLP-DNGTGSWSLSPVPGP----- 310
Db	262 QMKMLEIYNNKKVYSAELEYAAMKDDFVTLPKIDVGN--LDWTVIQRNDSAPIRHLD 318
QY	311 --APPLQTYPGQPRFSVQGSRYASSL-WTSSFGLAISGPRIFEDVRFQGERLVYEISIQ 366
Db	319 DRKSPRLTPEGRKRWAYGEEELYEFMSMDWGFYTSMSRDTGISFYDITFEGKERIYVELSIQ 378
QY	367 EALAIYGGNSPAAMTTRVVDGFGMGKXTTTLTRGVDCQYLATVYDWMH-FLEESQAPKI 425
Db	379 ELIATYGGSDDPENQHTTFYSDSYGVGNRFS-LVPGYDCATATGYFTTDFEYDEFENRKL 437
QY	426 RDAFCVFEONONGPLRRHNSDLYSHYFGGLAETVLVVRSMSTLLNVYDVMDFVHPHSGAI 485
Db	438 --SYCVFENQEDYSLLRHTGASYSAI---TQPTLVNRRISITIGNVDYFLKFFLDGL 492
QY	466 ELRFATATGYISSAFLPGAT-GKYGNOVSEHTLGYVHTSHAHKVDLDVAGLENTVAEDM 544
Db	493 EYSVRAAGYIAGYNNPETSAPYGLKIDVLSGSHDHVLANKVLDVDDGGTGNRRASKYVM 552
QY	545 VEVPAVAVWSPPEHQLORLQVTRKLEEMEE-----QAALFVGSATPRYILASNHS 594
Db	553 KDVDVEYEMAPQTVYNTNQIAREVLKEDDFNGINMPENOGGLLIESA-----EET 603
QY	595 NKMGPFRGRIQMLSPAGEPLPONS-SMARGFSWERYQLATVQRKEEPPSSSSVFENQNDP 653
Db	604 NSFGNPRAYNNIPGGGVYHRIKNSRSGPETONMARSNLETKHKHDEELRSSSTALTNTNL 663
QY	654 WAPYVDFSDFINNETIAGDILAAWYTAGLHLIPLHEDIINTTATVVGSGVGFPLRPYFFDP 713
Db	664 YDPRVNFNAFLDDESDIEDIYAWNLGLHLPLNSNDLPNTLFTSRHAASFMLTPRYFEDS 723
QY	714 DPSEYSAISYIFRGDOD 730

Db	724	ENSRDTTQGVFPYTDDE	740
RESULT 6			
Q920K6	PRELIMINARY;	PRT;	271 AA.
AC Q920K6			
DT 01-DEC-2001	(TREMBLrel. 19, Created)		
DT 01-DEC-2001	(TREMBLrel. 19, Last sequence update)		
DT 01-MAR-2002	(TREMBLrel. 20, Last annotation update)		
DE	Diamine oxidase (EC 1.4.3.6) (Fragment).		
OS	Cavia porcellus (Guinea pig).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
NC	Mammalia; Eutheria; Rodentia; Hystriocognathi; Caviidae; Cavia.		
NCBI_TaxID=10141;			
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-HARTLEY; TISSUE=SMALL INTESTINE;		
RT	Kitanaka J., Kitanaka N., Takemura M.;		
RL	"Cloning of a cDNA for guinea-pig diamine oxidase (histaminase)."		
DR	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AB073328; BAB/0498.1;		
DR	InterPro; IPR000269; CnH_oxidase.		
DR	Pfam; PF01119; Cu_amine_oxid; 1.		
DR	PROSITE; PS01165; COPPER_AMINE_OXID_2; UNKNOWN_1.		
KW	Oxidoreductase.		
FT	NON TER		
SO	SEQUENCE 271 AA; 31323 MW; CEC46D64E4AA135D CRC64;		
Query Match	15.0%; Score 615.5; DB 11; Length 271;		
Best Local Similarity	49.1%; Pred. No. 2.3e-42;		
Matches 115; Conservative 31; Mismatches 87; Indels 1; Gaps 1;			
Qy 491	ATGYSSAFLELGATGKYNQVSEHPLGTVHTPSAFKFDLDVAGLENNVMAEDVFPMA 550		
Db 1	ATGYHAFFPYPRGCGRTGLRLTHLGINIRHLHPRDMDVAGTKNSFQTLQMKLENT 60		
Qy 551	VPMSEHQLQRLQVTRKLEEEQAAFLVGSATPRYLVLASNHSKMGHPRGRIQMLSF 610		
Db 61	NPMSPGHLVQPTLKOSYSOERQAAFFGGTLPRYLLFTSRKRNMGHERGYRLQIISM 120		
Qy 611	AGEPLRQSSNARGSMWRQYLAQVQRKEEPESSSSVQNDPMAPTDQFSFI..NNETI 669		
Db 121	AEQVLPAGLPERRATIMARVPLAVTKRESELCSSSTIYNQNDPMQPPVFAQKFLHNEDI 180		
Qy 670	AGKDLVAVVTAGFLTHPAEDIPNTVTGNGVGFELRPYNFEDEDPSFYSADSI 723		
Db 181	ENEDLVAVVTGFLTHPHSEDPVNPATIGNSGFLFRFNFKEKDPSLASRTV 234		
RESULT 7			
Q9CWL6	PRELIMINARY;	PRT;	204 AA.
AC Q9CWL6			
DT 01-JUN-2001	(TREMBLrel. 17, Created)		
DT 01-JUN-2001	(TREMBLrel. 17, Last sequence update)		
DT 01-MAR-2002	(TREMBLrel. 20, Last annotation update)		
DE	1600012D06Rik protein (Fragment).		
GN	ABP1 OR 1600012D06Rik.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
NC	Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.		
NCBI_TaxID=10090;			
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=PLACENTA;		
RC	MEDLINE=21085660; Pubmed=11217851;		
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,		
RA	Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,		
RA	Aizawa K., Iwawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,		
RA	Saito T., Okazaki Y., Gojodori T., Bono H., Kasukawa T., Saito K.,		
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,		

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schirml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okada T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boftelli D., Bojunga N., Carninci P., de Bona M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gueniche S., Hill D., Hofmann M., Hume D.A., Kamlaya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.F.,
 RA Suzuki H., Togo-Oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszynski B., Yoshida K., Hasegawa Y., Kawaji H., Kohlschl S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection.",
 RT Nature 409:685-690(2001).
 RU EMBL: AK005423; BAB2401.1; -.
 DR MGI: 1923757; ADPL.
 DR InterPro: IPR000269; CuNH_oxidase.
 DR Pfam: PF01179; Cu_amine_oxid; 1.
 DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
 FT NON TER
 SQ SEQUENCE 204 AA; 23861 MW; 49PR7CBD6CD0555C CRC64;
 Query Match 11.5%; Score 472.5; DB 11; Length 204;
 Best Local Similarity 50.3%; Pred. No. 8.6e-31;
 Matches 95; Conservative 24; Mismatches 65; Indels 5; Gaps 4;
 Oy 572 EQQAAFLVGSATPRILYLASHNSKMGHPGRYRIQMLSTAGEPLDPONSSMARGFSWERYO 631
 Db 15 EHQAAFLVGSATPRILYLASHNSKMGHPGRYRIQMLSTAGEPLDPONSSMARGFSWERYO 631
 Oy 632 LAVTARKKEEPPSSSSVFNDNDPAPTVDESDFI--NNETAGDLYAVMTAGFLTHPHAD 690
 Db 75 LAVTARKKEEPPSSSSVFNDNDPAPTVDESDFI--NNETAGDLYAVMTAGFLTHPHAD 690
 Oy 691 IPRNTVYGVNGVGFLLRPYNEFDEPSFYSADSIYERGDODAGACEVNPACLPQAAACAP 750
 Db 135 VENTATPGNCVGFLLRPYNEFDEPSFYSADSIYERGDODAGACEVNPACLPQAAACAP 750
 Oy 751 DLPATSHGC 759
 Db 192 S--PPFSYNG 199
 RESULT 8
 Q9C6V7 PRELIMINARY; PRT; 741 AA.
 ID Q9C6V7
 AC Q9C6V7
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Copper amine oxidase (EC 1.4.3.6).
 GN F2/H3_13.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eudicots II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen K., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn J.E., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.O., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz G.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu X.X., Liu Z.A., Luros J.S., Malt R., Marshall A.,
 RA Miltner J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,

RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Tatum M.J., Town C.D.,
 RA Ullrich T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana".
 RT Nature 408:816-820(2000).
 CC -1- CATALYTIC ACTIVITY: KCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
 H(2)O(2).
 CC -1- COFACTOR: COPPER AND TOPAQUINONE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
 DR EMBL: AC074360; AAG60154.1; -.
 DR HSSP: 043077; 1KST.
 DR InterPro: IPR000269; CuNH_oxidase.
 DR Pfam: PF01179; Cu_amine_oxid; 1.
 DR Pfam: PF02727; Cu_amine_oxidn2; 1.
 DR Pfam: PF02728; Cu_amine_oxidn3; 1.
 DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
 DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
 KW COPPER_oxidoreductase, TPO.
 SQ SEQUENCE 741 AA; 84279 MW; DB69E786864B558 CRC64;
 Query Match 9.4%; Score 385; DB 10; Length 741;
 Best Local Similarity 25.3%; Pred. No. 9.2e-23;
 Matches 183; Conservative 97; Mismatches 280; Indels 164; Gaps 33;
 Oy 50 PTHPGQSLFADLSHEELTAVRFLTORLGLVD-----AAQAPSDNCVFS----- 98
 Db 104 PPPHP-----FDPLETEILNININERYPIGLEHRTFOYGLNEPKSLVSSVQ 158
 Oy 99 -VELQPPKAAALAHDRSPPARALIVFGRRPQGNVSELY-VGLPHPSYRVDY 156
 Db 159 YHNVSPPQAPVIAHDK--TRE--IVDPA--SQAVSKIVG----- 199
 Oy 157 VERHGGLPLHYHRPVLFOEYLDIDQMIENREL-----PASGLHHCCEYKRGRL--V 209
 Db 200 ---NGYPM-----LTIDEQATSELYLAKRPPRDS-----RRGLNVSEY 237
 Oy 210 TMTTAPRGLOSGDRATWEG-----LYNISGAGFLAHVGLVNHKRLDPA 257
 Db 238 VVTSITMG-----WFGKAPKRLIKRPFYLVNSVNTYLPIDGMIIIV--LDQM 286
 Oy 258 RMTIQKVFYQGRYDLSLAQLEAOFEGALNVVLIPINGTSGMS--LKSVPPEPAPLPQ 316
 Db 287 KVT-----FRDRETSPLPVA-----KTEFRISKLPKPGPSLQNAVLE 326
 Oy 317 YPOGPRFSYQGRVAVASSLMTF--SFGLA--FSGPRIFDVROGER-LVEISLOEALA 370
 Db 327 QSGPPEFKIDGHTNRANNEFHMSFDVRAGLVLSLSDMDVNRKROYLYKHLSEIV 386
 Oy 371 IYGGNSPAAMTTRYVD-GGFGMGKYYTPTLRGYDCYLAITYDMHFLLESQAQKTRDAF 429
 Db 387 PYMDPSEMDYFRFFFCGCGCOYAVSLPYDCGNAAFMGCVASPDGPIKKTNNW 446
 Oy 430 CVPEQNGQLPLRNHSDLSYHFGGLAETVLYVRSSTLLNDYVMDYVPHPGAEIIRF 489
 Db 447 CIFEKTAG--DIMKRTHEIIEPLGLKVRPDSLVVRKMTVVGNDYIVDYERKPGSKIGV 505
 Oy 490 YATG-----YISSAFEGATKXGNOVSEHTLCTVTHSAHFRVLDVADGLEMMVVA 541
 Db 506 GLTGVLVEVPKRVNTSEI-KEDDINGITVADNTIGVNDHFTYTRLDLDIDSTDSFVR 564
 Oy 542 EDWTF--VPMAYVWSPHOLQVTRKLLIEMEQAFLVGSATPRILYLASHNSKMGH 599
 Db 565 SELVTKRTPKSV-----NPKRSYMTTKLKAEE-----LVVNPBKRKYKHG 607
 Oy 600 PGYR-IGKLSFAGEPLPONSSMARGFSWERYOLAVTOKKEEPPSSVF--NQNDP-- 653
 Db 608 EYGRLLHGPASEGPLADDDYPOIRAFTNNVMTTPYNNTEVWASGLYADRSQDDDTL 667
 Oy 654 --WAPTVDESDFI--NNETAGDLYAVMTAGFLTHPHADIPNTVYGVNGVGFLLRPYNEF 711


```

OY 148 --HPSYMDVYVERHGGLPRHRYVLFQEYLDIDOMTFNRELPOASGLHHCCTYKRG 205
DB 134 HGFPSEFTELEFKASKLELTY--PPFKSLID-----RS 165
OY 206 RNLVMTTAPRGLOSGDRATWFG-----LYNISGAGFLHHY-GLELLVNHKA 253
DB 166 LNISVSCIP-----FTVGNKGETTTRRELKASCFYDGSNNVTRPEIGITITID--- 216
OY 254 LDPARMTIQKVFYQGRYDLSLAOLEAOFAGLVNVLLIPDNGTGSWSLK-SPVPGPAP 312
DB 217 VD---SMQVVKYSDRFKKPLPEKE-----GNDFRKKHKPF----- 248
OY 313 PLQFYPOGRPSVQSGSRVASSLSMTFSPGLGAFSSGRIFDV-----RFGQERLYEIS 364
DB 249 PPSGVNSTGTGKRIKGNRKAMKMHVGFATAGVITSTASVLDRTKFR--RVMYGQH 306
OY 365 LOELALAIYGSNSPAAMTTRYVD-GGFGMGKTYTPTLRGVDCPYLATYVDMHFLLESQAPK 423
DB 307 VSETFVPMDPTEMYETTFMDIGFGGRSAVNLQPLDQONAFDGHVAGPDGTAQ 366
OY 424 TIRDAFCVFEONQGLPRRHSDLY--SHYFGLAETVLVVRNSMTLLNTDYVMDTVF 480
DB 367 KATNMVCVFEKN-GYGASFRHTEINVPQVITSGAAEISLVVRVATLGNVDYIVDWEK 425
OY 481 PSGATEI-----RFATGYSASFLEGATGKYGNOVSEHTLCTVTHSAHFVDDY 532
DB 426 KSGALRVVDLTGLVEKATSTYSNEQI--TENYGTQVAKTIVNHDHYLYYLDLV 483
OY 533 AGLENMVAEDMVEFVPA--VPMSPHQLRLQVTRKLEMEQOAFVGSATPRYLYLAS 591
DB 484 DGNGLSLVAKAKTIVRYVDVKNKTSRKRSYVTVYKETAKEADGVRGLSEVELLYNP 543
OY 552 NNSKMGHPRGYRIOMLSFAGEPLPONS-----SNARGSWERYOLAVYORKEEBS 643
DB 544 NKKTIGTGVYRL-----IPEHLPATSLTDDYPEIAGTY--KYPVWV----- 588
OY 644 SSSVNONDPNAPRTYVDFSD-----FINNETIAGKDIYAVVATGFLIIPHAEDIP 692
DB 569 ---AYDKSERMAGGF-YSDRSRGDDGLAVWSRNRLEIKNDIWMYNGFHIIPOEDDP 644
OY 693 NTVTGNGVGFELRPYNFEDEDP 715
DB 645 VMPPLHG--GFTLRPSNFEEDNP 665

```

RESULT 11

097XMI PRELIMINARY; PRT; 660 AA.

AC 097XMI: PRT; 660 AA.

DT 01-OCT-2001 (TREMBLrel. 18, Created)

DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Amine oxidase (copper-containing) (tyna) (EC 1.4.3.6).

GN TYNA OR SSO1692.

OS Sulfolobus solfataricus.

OC Archaea: Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;

OC Sulfolobus.

OX NCBI_TaxID=2287;

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 35092 / DSM 1617 / P2;

RX MEDLINE=21332296; PubMed=11427726;

RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,

RA Aweyer M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,

RA De Moors A., Erasuo G., Fletcher C., Gordon P.M.K.,

RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,

RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,

RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,

RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.,

RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).

DR EMBL; AE006782; AAK41902.1; -

DR InterPro; IPR000269; Cunn_Oxidase.

```

DR Pfam; PF01179; Cu_amine_oxid. 1.
DR Pfam; PF02727; Cu_amine_oxid2; 1.
DR Pfam; PF02728; Cu_amine_oxid3; 1.
DR PROSITE; PS01164; COPPER AMINE OXID. 1; UNKNOWN. 1.
DR PROSITE; PS01165; COPPER AMINE OXID. 2; UNKNOWN. 1.
KW Oxidoreductase; Complete proteome
SQ SEQUENCE 660 AA; 77136 MW; 413121DF91C74076 CRC64;

```

Query Match 9.0%; Score 368.5; DB 17; Length 660;

Best Local Similarity 24.4%; Pred. No. 1.8e-21;

Matches 175; Conservative 103; Mismatches 303; Indels 137; Gaps 28;

```

OY 63 LSRELTVAMRPLTQRLPGVLDAQAARPSDNCVSEVLQLPKKAALAHLDGSPPPAR 122
DB 17 LNEEIRKSVEVLKROLN--LDAKVVK-----PFSVELKEKKQOEYLEWRNNKNIKIER 68
OY 123 EALATVFGROPQPVNSELVVGPLPHPSYMRDVTVERHGGPLPYRRPVLFOEYLDIDOM 182
DB 69 ESL-VKTYNFDOR-KYEAII-----SLDNNVKEINSIDAY--PRTTDEFECEKA 118
OY 183 IFNRELPOASGLHHCCTYKRGRLVMTTAPRGLOSGDRATWFGLYYNISGAGFLHH 242
DB 119 VANDKRVQEA-----LTKRGILINDNL--LWVDCWAPG---H 151
OY 243 VGEILLVNHKALDPAKMTIQKVFYQGRYDLSLAOLEAOFAGLVNVLLIPDNGTGSWSL 302
DB 152 VDEELRGRRVVAIG-YMW-VKQIEDNGYGRPHGLMPWVDLKMVEYIRIDHGT----- 203
OY 303 KSPVPGPRA-----PPLQF-YPGGPRFSVQSGSRVASSLSMTFSPGLGAFSG 346
DB 204 -SPLPLQDANITPEKLGITFGDGLKPIETROPLOSSIKINGEISIKYRRRLIGITPRRG 262
OY 347 PRIFDVRGGER-----LYTEISLOEALAIYGSNSPAAMTTRYVDG--FGMGKTYTPTLR 400
DB 263 LVIYDVRVIDENNRREMIYLRASVVDLVAVYDPSFHNKKVVLADAGVGLGNFTYPLSH 322
OY 401 G-----VDC-PYLATYVDMHFLLESQAPKTRDAFCVFEONQGLPRRHSDLYSHRG 453
DB 333 GWYDLNCDCEFEVYIHLDTVTVSSNGTPIKKIKKICVAHEDEFGVLMR--HFDLRS--- 376
OY 454 GLAET-----VLVVRNSMTLLNTDYVMDTVFHRSGAIEIFVATGYTISAFLE--GATGR 507
DB 377 GKSEVRRNRRLVVSFMATLANTDYGFPMYFYDDGSEFLYKLTGLTINDSISEKPTPKY 436
OY 508 GNOVSEHTLGTVTHSAHEKVDLVAGLENMVAEDMVEFVPAVPMSPHQLRLQVTRK 567
DB 437 GTRVTPVYAPRIHIFENRLNINVDGLNRIT-----EVNLRGEPITEK 481
OY 568 -----LMEEQOAFVGSATPRYLYLAS-NNSKMGHPRGYRIOMLSFAGEP 614
DB 482 NPGVNAFFAEENLLENEDARRHVPQGRYKIVNIQKKNLGLPVARLVPCHNVLP 541
OY 615 LPONSSMARGFSEMYROLAVTORKEEPPSSSVFQNDMPATYVDFSD-----FINNETI 669
DB 542 LPDDSVRRRGAVYINHLWVTPYNEERYASDY-----PYLRANDLPYTIILKKNSI 594
OY 670 ACKDLVAMVTAGFLIIPHAEDIPNTVTYVNGNGVGFELRPYNFEDEPSSFASDSIYFRG 727
DB 595 VDEDLVIMYTLGVEHVRIEDMP--VMPVEMAGFRILPGGFDKNNTIYLPQRLRING 650

```

RESULT 12

09UEU7 PRELIMINARY; PRT; 64 AA.

AC 09UEU7:

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Placenta copper monamine oxidase (Fragment).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA MEDLINE-96207227; PubMed-8619474;
RT "A 'double adaptor' method for improved shotgun library construction.";
RL Anal. Biochem. 236:107-113(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA MEDLINE-97264341; PubMed-9110174;
RT Yu W., Anderson B., Morley K.C., Muzny D.M., Ding Y., Liu W.,
  "Large-scale concatenation cDNA sequencing.";
RL Genome Res. 7:353-358(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA Yu W., Gibbs R.A.;
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF034985; AAC09346.1;
DR InterPro: IPR000269; CUNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
FT NON_TER
SQ SEQUENCE 64 AA; 6892 MW; A7A0F2972EAF3210 CRC64;

Query Match
Best Local Similarity 8.7%; Score 358; DB 4; Length 64;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 GGGFLLRPNFDEPFSADSIYRGPDGACVNPPLACPOAAACAPDLPARSHG 759
DB 1 GGGFLLRPNFDEPFSADSIYRGPDGACVNPPLACPOAAACAPDLPARSHG 60

QY 760 FSHN 763
DB 61 FSHN 64

RESULT 13
Q9ST12 PRELIMINARY; PRT; 756 AA.
AC 09ST12;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, last sequence update)
DE 01-MAR-2002 (TREMblrel. 20, last annotation update)
GN 14C9.130 OR AT4G12290.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID-3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Peters S.A., van Staveren M., Dirkse W., Stiekema W.,
  Bancroft I., Mewes H.W., Mayer K.F.X., Lemcke K., Mannhaupt G.,
  Scheller C.;
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Peters S.A., van Staveren M., Dirkse W., Stiekema W., Mewes H.W.,
  Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.

```

```

DR EMBL: AL080318; CAB45976.1; -
DR EMBL: AL161533; CAB78272.1; -
DR HSSP: Q43077; IKSI.
DR InterPro: IPR000269; CUNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
DR Pfam: PF02727; Cu_amine_oxid2; 1.
DR Pfam: PF02728; Cu_amine_oxid3; 1.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; UNKNOWN.1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
SQ SEQUENCE 756 AA; 84910 MW; 891019DEF0F649CA CRC64;

Query Match
Best Local Similarity 8.7%; Score 357.5; DB 10; Length 756;
Matches 163; Conservative 100; Mismatches 280; Indels 191; Gaps 29;

QY 94 NCVFSYELQLPKAAALAHLDGSPPARALAI-----VFEGROPQPNVSE 140
DB 126 HALHETVLEPERK-NLYRHEKGNPLPRKASVIARVAGDTVLTVDISTGVDNSNPV 184

QY 141 LVVG-PLPHSYMRDVTYRHHGRLPYHRRPVLFOEYDIDDMIFRELPOASGLLHHC 199
DB 185 RVSGYPMWTIEEKNDITV-----VFSNAD-----FKRTIIS----- 216

QY 200 FYKHGRNLVTMTTAPRGLOSGDRATWFG-----LYNISGAFELAHV-GLE 246
DB 217 ---RGVNLTDVIGCP--ISCG---WFGKKNENARVIKSGCFMGTGTFNFMPIEGLT 266

QY 247 LLYNKAALDPARNTICKYVQGRYDLSLAOLEAFAGLVNVLIPDNGTGSWSLSKSV 306
DB 267 ILID--LDTKQ-----VIEITDGRA-----IPI 288

QY 307 PEG-----PAPPIQF-YRQGRFSVQSGRYAS-SLMFSGLAGAFSP 347
DB 289 PGSTNDIYFQKLATTDKRPLNPISIEOPRGSPSEIENHLKMANWEHLKPPDRAY 348

QY 348 RIFDVR-----FOGERLYVEISLOEALAIYGGNSPAAATTRYVDG-FGSKYTTPLTR 400
DB 349 VISRVAVHPDHPHETRDVNYKGVSELFVPMYSDAWYFKTYMDAGEFGGLQAMPVLP 408

QY 401 GVDCLPLATYVDNHPLESQAPKTIIDAFVPEQN-----OGLPLRRHSD 447
DB 409 LNDGPRNAAVMDGVFAAADGTPEVRNENVCIFESYAGDIGRHSSEPIGPIREVR--- 465

QY 448 YSHFGLAETVLYVMSGSLTYDVMPTVFPSPGATIEIRFYATGYSASFGLATGY 507
DB 466 -----PKVTLVYMAASVGNITDIIDYEROTDGLIAKAGLSIL---WYKGTYYON 514

QY 508 GNOV-----SEHTIGTVHTSHAFKVDVDAVGLNMYVAEDVFEVPMAY 551
DB 515 KNOVERKDKNQNEELHGTLLSERVIGVINDHYVTFLDLDVQDPN-----SPVKYNLR 569

QY 552 PMSPEHQLOR---LOYTRKLEMEBOAFLVGSATP-RILYILASNSNKKMGPRGRION 607
DB 570 QETEPESPRKSYLKAVRINAKTEKDQIKLSLYDSEFEHVINSKTRVGNPTGKVPV 629

QY 608 LSPAGEPLPONSSMARGFERYOLAVTORKEEPESSSVFNONDMAPTVDFSPFINN 666
DB 630 RTTAAALILHDHPQKRGAFNTQNIWTPYKNSQAGLFTYQSGDDTLIATWSB--RD 687

QY 667 ETIAGDLVAVTAGFLIHPHAEDIPNTYVGVNGVFLRPYNFDEPFSADSIYR 726
DB 688 ROIENKDIYVWYTLGFHHIPQCEQDFIMPVSS--SFDLKPVNFFERNPLTSLAPNF--- 742

QY 727 GQDDAGACEVNP 740
DB 743 -EHDLPVCGVSVS 755

RESULT 14
Q9HBM7 PRELIMINARY; PRT; 194 AA.
AC 09HBM7;
DT 01-MAR-2001 (TREMblrel. 16, Created)

```


GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 07:58:08 ; Search time 84.2615 Seconds

(without alignments)
1578.232 Million cell updates/sec

Title: US-10-081-408-20

Sequence: 1 MDMLRNLFLMAAASINNA.....QAACAPDLPAFSGGFSHN 998

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq.101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3967	74.1	763	20	AAV03219 Amino acid sequence
2	2480.5	46.3	729	23	AAU084261 Human endometrial
3	1180	22.0	447	23	AAU10670 Mouse LICAM Fnd-5
4	1175.5	21.9	644	23	AAU10652 Chicken NG-CAM F80
5	1175	21.9	536	20	AAV21557 GS-6C6 fusion prot
6	1171.5	21.9	514	19	AAW47004 Glutathione-S-tran
7	1168	21.8	1252	22	AAE02008 Yeast cystathionin
8	1167.5	21.8	362	16	AAW79948 g17 antigen. Chlm
9	1166	21.8	658	21	AAV96056 Human suppressor o
10	1164.5	21.7	443	23	AAU10673 Chicken Nr-CAM Fnd

11	1164	21.7	285	13	AAE22141 GST: NABV 693:691
12	1164	21.7	442	23	AAU10667 Human LICAM Fnd-5
13	1163	21.7	1196	21	AAV58917 Fumonisin esterase
14	1163	21.7	1196	21	AAV68853 A fusion of aminop
15	1162.5	21.7	409	23	AAO17115 GlutathioneStransf
16	1162.5	21.7	409	23	AAW49711 Murine GST-modf fu
17	1162	21.7	315	13	AAE22138 CAP-N fusion prote
18	1162	21.7	361	13	AAE22145 GST: NABV 690:691
19	1161	21.7	272	19	AAW56327 Amino acid sequenc
20	1161	21.7	635	23	AAU10668 Mouse LICAM F80 fu
21	1160.5	21.7	258	20	AAW74117 GST/GI transport r
22	1160.5	21.7	277	20	AAW74116 GST/GI transport r
23	1160.5	21.7	298	23	AAU087086 Slglec-BMS-L3a cyt
24	1159.5	21.6	259	20	AAW74125 GST/GI transport r
25	1159.5	21.6	277	20	AAW74099 GST/GI transport r
26	1159.5	21.6	277	20	AAW74107 GST/GI transport r
27	1159.5	21.6	279	20	AAW74097 GST/GI transport r
28	1158.5	21.6	426	16	AAE87091 Turkey prolactin/G
29	1158	21.6	282	20	AAW74123 GST/GI transport r
30	1158	21.6	379	20	AAV31954 Glutathione-S-tran
31	1158	21.6	439	23	AAU10664 Chicken Ng-CAM Fnd
32	1157.5	21.6	282	20	AAW74095 Amino acid sequenc
33	1157.5	21.6	422	20	AAV23648 Chicken NG-CAM Fnd
34	1157.5	21.6	559	23	AAU10663 GST/GI transport r
35	1157.5	21.6	647	20	AAW73573 GST-Msp I methylas
36	1157	21.6	247	20	AAW74114 GST/GI transport r
37	1157	21.6	257	20	AAW74124 GST/GI transport r
38	1157	21.6	516	19	AAW47005 Glutathione-S-tran
39	1157	21.6	564	16	AAE84565 Trypanosoma cruzi
40	1157	21.6	643	16	AAE84568 Trypanosoma cruzi
41	1157	21.6	673	17	AAE98434 GST/human p57 fusl
42	1156.5	21.6	247	20	AAW74103 GST/GI transport r
43	1155.5	21.6	1140	16	AAE81731 GST-SEP fusion pro
44	1155.5	21.6	1140	23	AAE13360 pGEX-Sep45 plasmid
45	1155	21.6	259	20	AAW74130 GST/GI transport r

ALIGNMENTS

RESULT 1	AAV03219	standard: Protein; 763 AA.
ID	AAV03219:	
XX	21-JUN-1999	(first entry)
AC		
XX		
DE		Amino acid sequence of the vascular adhesion protein-1.
XX		
DE		Human: vascular adhesion protein-1; VAP-1; endothelial cell;
KW		Lymphocyte; inhibition; amine oxidase.
KM		
XX		
OS		Homo sapiens.
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 43	
FT	Misc-difference 47	/note- "potential O-glycosylation site"
FT	Misc-difference 47	/note- "potential O-glycosylation site"
FT	Misc-difference 679	/note- "potential O-glycosylation site"
FT	Misc-difference 679	/note- "potential O-glycosylation site"
FT	Misc-difference 137	/note- "potential O-glycosylation site"
FT	Misc-difference 137	/note- "potential O-glycosylation site"
FT	Misc-difference 232	/note- "potential N-glycosylation site"
FT	Misc-difference 232	/note- "potential N-glycosylation site"
FT	Misc-difference 294	/note- "potential N-glycosylation site"
FT	Misc-difference 294	/note- "potential N-glycosylation site"
FT	Misc-difference 592	/note- "potential N-glycosylation site"
FT	Misc-difference 592	/note- "potential N-glycosylation site"
FT	Misc-difference 618	/note- "potential N-glycosylation site"
FT	Misc-difference 618	/note- "potential N-glycosylation site"
FT	Misc-difference 666	/note- "potential N-glycosylation site"
FT	Misc-difference 666	/note- "potential N-glycosylation site"

PT Diagnosing endometrial cancer comprises determining expression of
 PT nucleic acid molecules or expression products that are differentially
 PT expressed in normal and malignant endometrium -

PS Claim 33; Page 131-134; 233pp; English.

CC The invention relates to diagnosing endometrial cancer in a subject
 CC suspected of having endometrial cancer comprising determining the
 CC expression of a set of nucleic acid molecules or expression products in
 CC an endometrial sample suspected of being cancerous, where the set of
 CC nucleic acid molecules comprises at least 2 nucleic acid molecules
 CC selected from 50 fully defined sequences as given in the specification.
 CC The nucleic acids are used as an array of at least 2 of the 50
 CC nucleic acids bound to a solid substrate. Also included is a solid-phase
 CC protein microarray comprising at least 2 antibodies or its antigen
 CC binding fragments, that specifically bind at least 2 different
 CC polypeptides from the 50 fully defined sequences as given in the
 CC specification, fixed to a solid substrate. The methods and arrays are
 CC useful for the diagnosis of endometrial cancer, selecting and monitoring
 CC treatment regimes and identification of lead compounds useful for the
 CC treatment of endometrial cancer. The present sequence is one of 50
 CC proteins differentially expressed between cancerous and non-cancerous
 CC samples.

CC Sequence 729 AA:

Query Match 46.3%; Score 2480.5; DB 23; Length 729;
 Best Local Similarity 64.1%; Pred. No. 4,1e-223;
 Matches 468; Conservative 86; Mismatches 145; Indels 31; Gaps 3;

OY 266 GPPSLPHPCPSVSPSAQPTHPGQSLFADLSREELTAVMRPTQRLGCGVDAAQARS 327
 DB 29 GSSSPCHPCPSVSHRQAPHPGQSLFADLSREELTAVMRPTQRLGCGVDAAQAPS 88
 OY 328 DNCVSEVELQPPKAAALHLDRGSPPPAREALAIYFEGROPQNSYLVGGLPPPSYM 387
 DB 89 DMCISVELQLPPKAAALHLDRGSPPPAREALAIYFEGROPQNSYLVGGLPPPSYM 148
 OY 388 RQVYVRRHGGLPYHRPVLFEQYLDIDOMIFRELPAQSGLLHCCFYKHRGRNLVTMT 447
 DB 149 RQVYVRRHGGLPYHRPVLFEQYLDIDOMIFRELPAQSGLLHCCFYKHRGRNLVTMT 206
 OY 448 TAPRGISGDRATWGLYNYNISAGFELHVGELLVNHKALDPAMTQKVPYQRYRD 507
 DB 207 ATPRGISGDRATWGLYNYNISAGFELHVGELLVNHKALDPAMTQKVPYQRYRD 566
 OY 508 SLAOLEAOPFACLVNVLIPDNGTGSWSLSPVPPAPLOFYQGRFVSQGRVYS 567
 DB 267 DLGQLEKREKSRLEVRPPLPPNCASSLRNSPGLPPLDFOFSGQSYQGNLTVYS 326
 OY 568 SLWTSFSGCAGSPRIEDVRFQGERLYEISLQALAIYCGNSPAMTTRYVGGFGMG 627
 DB 327 SLWTSFSGCAGSPRIEDVRFQGERLYEISLQALAIYCGNSPAMTTRYVGGFGMG 686
 OY 628 KTTPLTRGVDCPYLATYVDWHFLESQAPKTRDAFCVFEONOGAPLRHSHDLXSHF 667
 DB 387 RNSRGLVRGVDCPYLATYVDWHFLESQAPKTRDAFCVFEONOGAPLRHSHDLXSHF 446
 OY 688 GGLAETVLVVRSMSTLANTDYVDVFFHPSGAEIRFYATGYSSAFLEGATG--XYGNO 745
 DB 447 GGLAETVLVVRSMSTLANTDYVDVFFHPSGAEIRFYATGYSSAFLEGATG--XYGNO 804
 OY 746 VSEHTLGVHTSHAEKVDLVAAGLENNWVAEDMVFVPAVWSPSEHOLQRLQVTRKLE 805
 DB 507 VSEHTLGVHTSHAEKVDLVAAGLENNWVAEDMVFVPAVWSPSEHOLQRLQVTRKLE 864
 OY 806 MEDQAFLVGSATPRYLTLASNSKNWGPGRYRIOMLSFAGPLPQONSSMARGESWERY 865
 DB 567 KEDLTAFFSLSGSLPRLYLYLASNOTNAMGHQRG-----Y 599
 OY 866 OLAVTQKKEEPSSSVFNPONDPWAPTVDFSDFINNETTAGKDLVAMVTAGFLHHAED 925
 DB 600 OLAVTQKKEEPSSSVFNPONDPWAPTVDFSDFINNETTAGKDLVAMVTAGFLHHAED 984

OY 926 INTVTGVNGVGFELRPYNFEDEDPSEYSGADSYFRGDODAGACEVNPACLPQAAACAP 985
 DB 660 INTVTGVNGVGFELRPYNFEDEDPSEYSGADSYFRGDODAGACEVNPACLPQAAACAP 719

OY 986 DLPAFSGHGF 995
 DB 720 DLPAFSGHGF 729

RESULT 3

AAU10670
 ID AAU10670 standard; Protein: 447 AA.

AAU10670;

14-FEB-2002 (first entry)

Mouse LICAM Fn4-5 fusion polypeptide.

Neurite outgrowth; fibronectin type III repeat; cell adhesion molecule;

Fn4-5; neurone; peripheral nerve damage; trauma; gliosis;

infection; degenerative disease; malignant disease; antibacterial;

central nervous system lesion; virucide; antiparkinsonian; nocotropic;

neuroprotective; antiinflammatory; mouse; LICAM; mutant; mutain.

Mus sp.

Synthetic.

US6313265-B1.

24-JUL-1995; 95US-0506296.

24-JUL-1995; 95US-0506296.

(SCRI) SCRIPPS RES INST.

Phillips G, Cunningham BA, Crossin KL;

WPI: 2002-017011/02.

Claim 3; Fig 25; 132pp; English.

The present invention relates to polypeptides that promote neurite
 growth. The polypeptides contain fibronectin type III repeats derived
 from a family of cell adhesion molecules (CAMs). The polypeptides of the
 invention include the F80, 3-5, and 4-5 regions of the CAM family
 members chicken N9-CAM, chicken Nt-CAM, mouse LICAM and human LICAM.
 The polypeptides of the invention are useful for promoting neurite
 outgrowth of neuronal cells in vitro e.g. in a cell culture system, or
 in vivo for treating disorders such as peripheral nerve damage
 associated with physical or surgical trauma, infarction, bacterial or
 viral infections, toxin exposure, degenerative disease, malignant
 disease that affects peripheral or central neurones, or in surgical or
 transplantation methods in which new neuronal cells from brain, spinal
 cord or dorsal root ganglia are introduced and require stimulation of
 neurite outgrowth from the implant and innervation into the recipient
 tissue, where the diseases include central nervous systems lesions,
 gliosis, Parkinson's disease, Alzheimer's disease, gliotic response or
 inflammation. The present sequence represents mouse LICAM Fn4-5 fusion
 polypeptide.

Sequence 447 AA:

Query Match 22.0%; Score 1180; DB 23; Length 447;
 Best Local Similarity 69.7%; Pred. No. 2e-101;

Matches 251; Conservative 9; Mismatches 50; Indels 50; Gaps 10;

```

OY 38 SPIIGYWKIKGLVOPTRLLLEYLEEKYEHLIERDEGKMKKKFELGLEPNNLPYYIDG 97
   |||||||
Db 2 SPIIGYWKIKGLVOPTRLLLEYLEEKYEHLIERDEGKMKKKFELGLEPNNLPYYIDG 61
OY 98 DVKLTQSMATIRIYADKKNMGLGSPKERAISMLEGAVLDIRGVSRAYSKEFETLKVD 157
   |||||||
Db 62 DVKLTQSMATIRIYADKKNMGLGSPKERAISMLEGAVLDIRGVSRAYSKEFETLKVD 121
OY 158 FLTKLPEMLKMFEDRLSKRTYLNLDHVTHTPDMLYDALDVVLYMDPKLDAFPKLVSEFK 217
   |||||||
Db 122 FLTKLPEMLKMFEDRLSKRTYLNLDHVTHTPDMLYDALDVVLYMDPKLDAFPKLVSEFK 181
OY 218 RIEAIPQIDKYLKSSKYIAMPLOGMOATFEGGDHPKQSQSEALFQGGDGEPSQLP-HC 276
   |||||||
Db 182 RIEAIPQIDKYLKSSKYIAMPLOGMOATFEGGDHPKQSQSEALFQGGDGEPSQLP-HC 237
OY 277 PSVSPSAQPT-----HP--GQS--QLFADLSREELAVMRFLTORLGP 316
   |||||||
Db 238 QSDTSLHLHMQPRLSHNGVLTYLSTYHPVEGSEKEQLFNLSDDELRT--HNLT-NLNP 294
OY 317 GLVMAQAARPSDNCVFSELDLPKAAALAHLDGSPPPA--REALIAVFFGRQPPNVS 374
   |||||||
Db 295 DL-----QYRFOLO-----AVTQOGGPGQAIIVREGGTALFGKPDGNGIS 334

```

RESULT 4

AAU10662
ID AAU10662 standard; Protein: 644 AA.

AC AAU10662;

DT 14-FEB-2002 (first entry)

XX Chicken Ng-CAM F80 fusion polypeptide.

XX Neurite outgrowth; fibronectin Type III repeat; cell adhesion molecule;

XX F80; neurone; peripheral nerve damage; trauma; gliosis;

XX Infarction; degenerative disease; malignant disease; antibacterial;

XX Central nervous system lesion; viricide; antiparkinsonian; nootropic;

XX Neuroprotective; antiinflammatory; chicken; Ng-CAM; mutant; mutein.

XX Gallus sp.

OS Synthetic.

XX US6313265-B1.

PN 06-NOV-2001.

PD 24-JUL-1995; 95US-0506296.

PF 24-JUL-1995; 95US-0506296.

XX 24-JUL-1995; 95US-0506296.

PR (SCRT) SCRIPPS RES INST.

XX Phillips G, Cunningham BA, Crossin KL;

PI WPI; 2002-017011/02.

DR Polypeptide for promoting neurite out-growth useful for treating

XX diseases such as inflammation, Parkinson's disease, trauma, comprises

PT fibronectin type III repeats derived from a family of cell adhesion

XX molecules -

PS Example 2; Fig 17; 132pp; English.

XX The present invention relates to polypeptides that promote neurite

CC growth. The polypeptides contain fibronectin Type III repeats derived

CC from a family of cell adhesion molecules (CAMs). The polypeptides of the

CC invention include the F80, 3-5, and 4-5 regions of the CAM family

CC members chicken Ng-CAM, chicken Nr-CAM, mouse L1CAM and human L1CAM.

CC The polypeptides of the invention are useful for promoting neurite

CC outgrowth of neuronal cells in vitro e.g. in a cell culture system, or
CC in vivo for treating disorders such as peripheral nerve damage
CC associated with physical or surgical trauma, infarction, bacterial or
CC viral infections, toxin exposure, degenerative disease, malignant
CC disease that affects peripheral or central neurones, or in surgical or
CC transplantation methods in which new neuronal cells from brain, spinal
CC cord or dorsal root ganglia are introduced and require stimulation of
CC neurite outgrowth from the implant and innervation into the recipient
CC tissue, where the diseases include central nervous systems lesions,
CC gliosis, Parkinson's disease, Alzheimer's disease, gliotic response or
CC inflammation. The present sequence represents chicken Ng-CAM F80 fusion
CC polypeptide.

SO Sequence 644 AA;

Query Match 21.9%; Score 1175.5; DB 23; Length 644;

Best Local Similarity 63.5%; Pred. No. 9.5e-101;

Matches 250; Conservative 19; Mismatches 52; Indels 73; Gaps 10;

```

OY 38 SPIIGYWKIKGLVOPTRLLLEYLEEKYEHLIERDEGKMKKKFELGLEPNNLPYYIDG 97
   |||||||
Db 2 SPIIGYWKIKGLVOPTRLLLEYLEEKYEHLIERDEGKMKKKFELGLEPNNLPYYIDG 61
OY 98 DVKLTQSMATIRIYADKKNMGLGSPKERAISMLEGAVLDIRGVSRAYSKEFETLKVD 157
   |||||||
Db 62 DVKLTQSMATIRIYADKKNMGLGSPKERAISMLEGAVLDIRGVSRAYSKEFETLKVD 121
OY 158 FLTKLPEMLKMFEDRLSKRTYLNLDHVTHTPDMLYDALDVVLYMDPKLDAFPKLVSEFK 217
   |||||||
Db 122 FLTKLPEMLKMFEDRLSKRTYLNLDHVTHTPDMLYDALDVVLYMDPKLDAFPKLVSEFK 181
OY 218 RIEAIPQIDKYLKSSKYIAMPLOGMOATFEGGDHPKQSQSEALFQGGDGEPSQLP-HC 277
   |||||||
Db 182 RIEAIPQIDKYLKSSKYIAMPLOGMOATFEGGDHPKQSQSEALFQGGDGEPSQLP-HC 238
OY 278 SVSP-----SAQPTHPGQSOLFADLSREELAVMRFLTORLGP 317
   |||||||
Db 229 AEDPPFPFVALTYGGDARGALLGLRPM-----SRQL-RVLVFNRGSDP- 284
OY 318 LVDAQAARPSDNCVFSE-LQLPKAAALAHLDGSPPPAREALATY---FFGR----- 367
   |||||||
Db 285 -----PSEPIAFETPEGVGPPPEELRVERLD-----DTLVSVERRTFKRSITGY 329
OY 368 -----QPQNVSELVVGPLPHSYMRDVTVERHG 396
   |||||||
Db 330 VLRYQVVERG-----SALPGSVLRDPQCDLRG 357

```

RESULT 5

AAAY21557
ID AAY21557 standard; Protein: 536 AA.

AC AAY21557;

DT 03-AUG-1999 (first entry)

XX GS-GC6 fusion protein.

XX Cellular senescence; modulator; GC6 gene; senescent gene expression;

XX pGC6; human; fusion protein.

XX Homo sapiens.

XX WO9925878-A2.

PN 27-MAY-1999.

PD 19-NOV-1998; 98WO-US24996.

PF 19-NOV-1997; 97US-0974180.

XX (GERO-) GERON CORP.

PI Funk W:
 XX WPI: 1999-347496/29.
 DR N-PSDB: AAX77118.
 XX
 PT New human GC6 gene, useful for identifying agents for treating
 XX diseases and/or conditions associated with cell senescence
 PS Disclosure: Page 29-30; 79pp: English.

CC The invention relates to methods for modulating and identifying cellular
 CC senescence. Recombinant expression vectors comprising a recombinant
 CC polynucleotide corresponding to a polynucleotide in a human GC6 gene, are
 CC useful for altering senescent gene expression. The vectors and host cells
 CC comprising the vectors are useful for identifying agents that prevent or
 CC modulate senescent gene expression. The polynucleotides are useful for
 CC producing the protein, pGC6 and nucleic acid derivatives. The proteins
 CC encoded are useful for raising antibodies specific for pGC6, which are
 CC useful for isolating pGC6, and for detecting cells comprising pGC6 in
 CC complex cell mixtures. The characterization of the polynucleotides enable
 CC the identification of therapeutic agents that identify and distinguish
 CC between young and senescent cells. This enables treatment of aging
 CC diseases induced or exacerbated by cellular senescence.

XX Sequence 536 AA:

Query Match 21.9% Score 1175; DB 20; Length 536;
 Best Local Similarity 49.6%; Pred. No. 7.9e-101;
 Matches 279; Conservative 41; Mismatches 123; Indels 120; Gaps 19;

QY 38 STILGWMKKGLOVPRLLLEKEEYERDEGDKMRNKEFLGEPNLPYIDG 97
 DB 2 STILGWMKKGLOVPRLLLEKEEYERDEGDKMRNKEFLGEPNLPYIDG 61
 QY 98 DVKLTOSMAIIRYIADKNNMLGSGPERAEISLMEGAVDIRGVSRIVSKDFETLYD 157
 DB 62 DVKLTOSMAIIRYIADKNNMLGSGPERAEISLMEGAVDIRGVSRIVSKDFETLYD 121
 QY 158 FLTKLPEMLKMFEDRLSHNTYINGDHYTHPDFMLYDALDVLVMDMCLDAEPKLVSPK 217
 DB 122 FLTKLPEMLKMFEDRLCHNTYINGDHYTHPDFMLYDALDVLVMDMCLDAEPKLVSPK 181
 QY 218 RLEAIPQIDKYLKSSKYTIAMPLOGQATFGGDDHPKSSQSEALFQGGGGEPSQLPHCP 277
 DB 182 RLEAIPQIDKYLKSSKYTIAMPLOGQATFGGDDHPKSPD---LIEG--RGIPRNSARBD 235
 QY 278 SVSPSAQPTWHPGQSLFADLSREELTAVMRFLTORLGGV----- 319
 DB 236 YFTNANRELKDAQODYHLEYVAMENSTHTTIEFTRELHMCIDINDKSIDTSTVAVIAYHN 295
 QY 320 -DAQAARP-----SDNCVFSEVLEQLPPKAALAHLDGRSPPARALATVEFGROPQPNYS 374
 DB 296 EDAGEGEPYHDSNRGTSKRLRLNPEKTSVL-----STALYF----- 333
 QY 375 ELVGPRLPSPSYMRDYTVERRHGRPLRYHRRVPLFOE---YLDIDOMIFRRELPRQASGLH 431
 DB 334 DLVNDVPIPN--KDTTYCOMFKLP-----VFOEKHNVIKVEPIYIGSH---ESLVH 381
 QY 432 HCCFYKRGKRNLYVTMTARGLQSGDRA-----TWPGLYYNISAGNF- 473
 DB 382 HLLLYO-----CSNNFNDVLESHECHPNMPDAFLTCETVIFAM-----AIGEGFS 430
 QY 474 FLHNVGELLVNHKALDPARMTIOKFYOGRYUDS-----LAOLEAQFEGLVYN 522
 DB 431 YRPVHGSLP---GTPLDP--HYVLELVHYDNPTEYEGSLDINSGLRLFTYMDIKRYDGVLE 486
 QY 523 VVLIDNGTGGSM-SLKSVPSPG 544
 DB 487 -----AGLMVSLFHTIPPG 500

RESULT 6
 AAW47004

ID AAW47004 standard; Protein: 514 AA.
 XX AAW47004;
 AC 13-AUG-1998 (first entry)
 XX
 DT Glutathione-S-transferase and hTERT fusion protein 4.
 XX
 DE Human; telomerase reverse transcriptase; hTERT; TRT; diagnosis;
 XX prognosis; cell proliferation; cancer; ageing; ribonucleoprotein.
 XX Homo sapiens.
 XX
 Key Location/Qualifiers
 FT Region 1..220 "glutathione-S-transferase fragment"
 FT Region 237..514
 FT Region /note="hTERT protein fragment"
 PN GB2317891-A.
 PD 08-APR-1998.
 XX
 PF 01-OCT-1997; 97GB-0020890.
 XX
 PR 14-AUG-1997; 97US-0915503.
 PR 01-OCT-1996; 96US-0724643.
 PR 18-APR-1997; 97US-0844419.
 PR 25-APR-1997; 97US-0846017.
 PR 06-MAY-1997; 97US-0851843.
 PR 09-MAY-1997; 97US-0854050.
 PR 14-AUG-1997; 97US-0911312.
 PR 14-AUG-1997; 97US-0912951.
 XX
 PA (GERO-) GERON CORP.
 PA (UYTE-) UNIV. TECHNOLOGY CORP.
 PI Andrews WH, Cech TR, Chapman KB, Harley C, Lingner J;
 PI Morin GB, Nakamura T, Harley CB;
 XX WPI: 1998-171633/16.
 XX
 PS pure and recombinant human Telomerase Reverse Transcriptase and its
 XX variants - are useful in the diagnosis, prognosis and treatment of
 XX cell proliferation conditions especially cancer and ageing
 XX
 PS Example 6; Page 226-227; 387pp: English.

The present sequence represents a fusion protein from an example of the present invention which describes human telomerase reverse transcriptase (hTERT). The present invention also describes the following methods: (A) determining whether a test compound is a modulator of hTERT, by detecting the change in hTERT recombinant protein or polynucleotide, on administration of the compound; (B) preparation of recombinant telomerase by contacting a protein preparation of hTERT with a telomerase RNA component; (C) detection of the hTERT RNA or protein in a sample by binding a relevant probe to the sample and detecting the complex formed or in the case of RNA detection, amplifying the product and correlating the presence of complex or amplification product with presence of hTERT in the sample; and (D) increasing the proliferation of a vertebrate cell by increasing hTERT expression; and (E) the use of an agent that causes an increase in cell vertebrate cell proliferation to create a medicament that inhibits ageing. A protein preparation of hTERT and the polynucleotide encoding hTERT can be used in the manufacture of medicaments for inhibiting the effect of ageing or cancer. Inhibitors of telomerase activity can be used to treat conditions that are associated with high telomerase activity. A protein preparation of hTERT can also be used in the new methods.

Sequence 514 AA:

```

Query Match      21.9%; Score 1171.5; DB 19; Length 514;
Best Local Similarity 75.4%; Pred. No. 1.6e-100;
Matches 236; Conservative 7; Mismatches 39; Indels 31; Gaps 5;

OY 38 SPILGYWKIKGLVOPTRLLLEYLEEKEYEHLIERDEGKMKRKKFELGEPNLPYITDG 97
   |||||
DB 2 SPILGYWKIKGLVOPTRLLLEYLEEKEYEHLIERDEGKMKRKKFELGEPNLPYITDG 61
   |||||
OY 98 DVKLTQSMATIRYIADKHNMLGSPKERAETSMLEGAVIDIRYGVSRAYSKEFTLKVD 157
   |||||
DB 62 DVKLTQSMATIRYIADKHNMLGSPKERAETSMLEGAVIDIRYGVSRAYSKEFTLKVD 121
   |||||
OY 158 FLSLKPEMLKMFEDRLSHKTYLNGDVTNPDPMLYDALDVLYMDPKCLDAFPKLVSEFK 217
   |||||
DB 122 FLSLKPEMLKMFEDRLCHKTYLNGDVTNPDPMLYDALDVLYMDPKCLDAFPKLVSEFK 181
   |||||
OY 218 RIEALPQIDKILKSSKYIAMPLOGQATFGGDDHPKPSDLYPRGSRASVGSVTKATISLE 241
   |||||
DB 182 RIEALPQIDKILKSSKYIAMPLOGQATFGGDDHPKPSDLYPRGSRASVGSVTKATISLE 241
   |||||
OY 260 ALFQGDGDEPS--QLPHC--PSVSPSAQPTWTHP-----GQSOLFADLSREELTAVMR 308
   |||||
DB 242 GALSGTRHSHSVGRONHAGPSPSRPRPMDTPCPPIVYATKHTLYSSGDKQLRP--S 299
   |||||
OY 309 FLTRQLGPGLVDA 321
   |||||
DB 300 FLTSLRPSLTGA 312
   |||||

RESULT 7
AAE02008
ID AAE02008 standard; Protein: 1252 AA.
XX
XX AAE02008;
AC
XX
XX 31-JUL-2001 (first entry)
DT
XX
DE Yeast cystathionine beta-synthase (CBS).
XX
KW Yeast; enzymatic cycling assay; cystathionine beta-synthase; CBS;
KM cystathionine beta-lyase; CBL; cyclical interconversion;
KM homocysteine.
OS Saccharomycetes cerevisiae.
XX
XX Key Location/Qualifiers
FH 1..238
FT /label= "Glutathione_S-transferase"
FT /note= "GST fused to the amino-terminal of CBS"
FT Region 239..745
   /label= "CBS-enzyme"
   /note= "Yeast cystathionine beta-synthase"
FT MISC-difference 459
   /note= "Encoded by GAG"
FT MISC-difference 460
   /note= "Encoded by GGA"
FT MISC-difference 531
   /note= "Encoded by ACC"
FT Region 746..1252
   /note= "Repeated CBS region"
XX
XX MO200133187-A2.
XX
XX 10-MAY-2001.
XX
XX 02-NOV-2000; 2000WO-US30518.
XX
XX 02-NOV-1999; 99US-0163126.
XX 10-MAY-2000; 2000US-0203349.
XX 01-NOV-2000; 2000US-0704036.
XX
XX (CATC-) CATCH INC.
XX

```

```

PI Kawasaki G, Webb HK, Owens J, Liedtke R, Forest D, Legaz M;
PI Lawson S;
XX
XX WPI: 2001-335847/35.
DR N-PSDB: AAD06034.
XX
PT Enzyme assays comprising using amounts of e.g. pyruvate and ammonia
PT produced during cyclical interconversion of homocysteine and
PT cystathionine to assess amounts of homocysteine and/or cystathionine in
PT a sample
XX
XX Claim 103; Page 75-79; 80pp; English.
PS
XX The present invention relates to an enzymatic cycling assays using
XX cystathionine beta-synthase (CBS) and cystathionine beta-lyase (CBL) to
XX assess the amount of homocysteine and/or cystathionine in a sample by
XX assaying the amount of pyruvate, ammonia, and/or nicotinamide adenine
XX dinucleotide (NAD+) produced during the cyclical interconversion of
XX homocysteine and cystathionine. The homocysteine reacts with L-serine to
XX form cystathionine in the presence of the enzyme CBS. The CBL catalyze
XX the conversion of cystathionine to homocysteine, pyruvate and ammonia.
XX The enzymatic cycling assay is less expensive and provides a higher
XX sample throughput than the diagnostic assays currently available.
XX The present sequence is yeast CBS having an amino terminal GST
XX fusion protein attached as a result of cloning into the bacterial
XX expression vector pGEX6P-2.
XX
XX Sequence 1252 AA:

Query Match      21.8%; Score 1168; DB 22; Length 1252;
Best Local Similarity 76.2%; Pred. No. 1.4e-99;
Matches 234; Conservative 12; Mismatches 37; Indels 24; Gaps 4;

OY 38 SPILGYWKIKGLVOPTRLLLEYLEEKEYEHLIERDEGKMKRKKFELGEPNLPYITDG 97
   |||||
DB 2 SPILGYWKIKGLVOPTRLLLEYLEEKEYEHLIERDEGKMKRKKFELGEPNLPYITDG 61
   |||||
OY 98 DVKLTQSMATIRYIADKHNMLGSPKERAETSMLEGAVIDIRYGVSRAYSKEFTLKVD 157
   |||||
DB 62 DVKLTQSMATIRYIADKHNMLGSPKERAETSMLEGAVIDIRYGVSRAYSKEFTLKVD 121
   |||||
OY 158 FLSLKPEMLKMFEDRLSHKTYLNGDVTNPDPMLYDALDVLYMDPKCLDAFPKLVSEFK 217
   |||||
DB 122 FLSLKPEMLKMFEDRLCHKTYLNGDVTNPDPMLYDALDVLYMDPKCLDAFPKLVSEFK 181
   |||||
OY 218 RIEALPQIDKILKSSKYIAMPLOGQATFGGDDHPKPSDLYPRGSRASVGSVTKATISLE 241
   |||||
DB 182 RIEALPQIDKILKSSKYIAMPLOGQATFGGDDHPKPSDLYPRGSRASVGSVTKATISLE 240
   |||||
OY 273 -----LPHCPVSPSAQPTWTHPQGSQLPADLSREELTAVMRFLTQRLPG 317
   |||||
DB 241 KSEGOADSRHNVLDVGTWTPILALKLPKALGIRKQIYAKL---ELYNPGSINRIRAKS 297
   |||||
OY 318 LVDAQA 324
   |||||
DB 298 WVEEA 304
   |||||

RESULT 8
AAR79948
ID AAR79948 standard; Protein: 362 AA.
XX
XX AAR79948;
AC
XX
XX 26-FEB-1996 (first entry)
DT
XX
DE G17 antigen.
XX
XX Syphilis; G15 antigen; membrane antigen; glutathione-S-transferase;
XX GST; fusion protein; membrane antigen; diagnosis.
XX
XX Chimeric Treponema pallidum;.
OS Chimeric unidentified bacterium.

```

```

XX Key Location/Qualifiers
FH Protein 1..228
FT /label= GST
FT Protein 229..362
FT /label= 17k-antigen
XX
XX EP670494-A2.
XX
XX 06-SEP-1995.
XX
XX 28-FEB-1995; 95EP-0400423.
XX
XX 28-FEB-1994; 94JP-0054672.
XX
XX (FURE ) FUJI REBIO INC.
XX (FURE ) FUJI REBIO KK.
XX
XX Fujii N, Fujimura K, Okada M, Ueno E;
XX
XX WPI; 1995-303656/40.
XX
XX New assay for anti-Treponema pallidum antibody - useful for
XX diagnosis of syphilis.
XX
XX Disclosure; Page 10-11; 18pp; English.
XX
XX G17 antigen (AAR79948) is a fusion protein composed of glutathione-
XX S-transferase and the 17 kDa membrane antigen of Treponema
XX pallidum. It is obtd. by insertion of a DNA fragment coding for
XX the 17 kDa antigen into vector pWGA and expression in E. coli.
XX G17 and similarly produced G15 (AAR79947) are used for diagnosis
XX of syphilis.
XX
XX Sequence 362 AA:
SQ
Query Match 21.8%; Score 1167.5; DB 16; Length 362;
Best Local Similarity 70.2%; Pred. No. 2.1e-100;
Matches 245; Conservative 15; Mismatches 56; Indels 33; Gaps 7;
QY 38 SPILGWKIKGLVOPTRLLLEYLEEKYEHLRYERDEGDKRNKKFELGEPNLPYYIDG 97
DB 2 SPILGWKIKGLVOPTRLLLEYLEEKYEHLRYERDEGDKRNKKFELGEPNLPYYIDG 61
QY 98 DVKLTQSMALIRYIADKHNMLGSGPKERAEISMLEGAVLDIRGVSRIVASKDFETLKVD 157
DB 62 DVKLTQSMALIRYIADKHNMLGSGPKERAEISMLEGAVLDIRGVSRIVASKDFETLKVD 121
QY 158 FLSKLPMLKMFEDRLSHKTYTLNGDHYTHDPMLYDALDVLYLMDPCLDAPFKLYSFKK 217
DB 122 FLSKLPMLKMFEDRLCHKTYTLNGDHYTHDPMLYDALDVLYLMDPCLDAPFKLYCFKK 181
QY 218 RIRAIPOIDKYLLSKSYIAMPLOGWQATFGGSDHPKPSQSEALFQGGDGEPSQLPHCP 277
DB 182 RIRAIPOIDKYLLSKSYIAMPLOGWQATFGGSDHPKPSD---LVPRG-----SELVSC 222
QY 278 SVSPSAQPTWHPGOSQ-----LFADLSREELTAVARFETLORLGPGLVDAQAAPSDNC 330
DB 233 TVCP-----HAGKAKAEKVECALKGIFRGITLPAD-----CPG-IDITVTFNADGT 278
QY 331 VFSVEYQLPDKAAALAHLDGSPPPAREA--LAIVFEGQPOPANVSEL 376
DB 279 AQRVELALEKKSAPSPLYRGTMVREDGIVELSLVSSEOSKAPHEKEL 327

```

```

DE Human suppressor of fused, hSu(fu), GST fusion.
XX
XX Suppressor of fused; hSu(fu); PRO1280; human; hedgehog;
XX signal transduction; cancer; tumour suppressor; therapy; diagnosis;
XX glutathione S-transferase; GST.
XX
XX Chimeric - Homo sapiens.
XX Chimeric - Schistosoma sp.
XX
XX Key Location/Qualifiers
FH Protein 1..226
FT /label= GST
FT Protein 227..658
FT /label= hSu(fu)
XX
XX WO200052166-A1.
XX
XX 08-SEP-2000.
XX
XX 02-MAR-2000; 2000WO-US05746.
XX
XX 05-MAR-1999; 99US-0123090.
XX 25-MAY-1999; 99US-0135736.
XX
XX (GETH ) GENENTECH INC.
XX
XX De Sauvage FJ, Gurney AL, Murone M, Rosenthal A, Stone DM;
XX Wood WJ.
XX
XX WPI; 2000-565502/52.
XX
XX New human suppressor of fused proteins and nucleic acids encoding them
XX useful e.g. as hybridization probes, in chromosome and gene mapping, in
XX the generation of antisense RNA and DNA
XX
XX Example 6; Fig 11; 67pp; English.
XX
XX The present sequence is that of a fusion between glutathione
XX S-transferase (GST) and human suppressor of fused (hSu(fu), see
XX also AA96054), a protein that functions as a negative regulator of
XX the hedgehog segment polarity pathway. hSu(fu) is highly expressed
XX in regions of active cell proliferation, and is probably a tumour
XX suppressor protein. hSu(fu) was expressed as a GST fusion protein
XX using vector pGEX in transformed bacterial cells, and used to study
XX the protein-protein interactions of hSu(fu). The invention provides
XX hSu(fu) nucleic acids (see AA50501), proteins, host cells, vectors,
XX antibodies, agonists and antagonists, as well as a method of
XX determining whether a disorder is modulated by hedgehog signalling.
XX
XX Sequence 658 AA:
SQ
Query Match 21.8%; Score 1166; DB 21; Length 658;
Best Local Similarity 87.2%; Pred. No. 7.7e-100;
Matches 225; Conservative 6; Mismatches 23; Indels 4; Gaps 2;
QY 38 SPILGWKIKGLVOPTRLLLEYLEEKYEHLRYERDEGDKRNKKFELGEPNLPYYIDG 97
DB 2 SPILGWKIKGLVOPTRLLLEYLEEKYEHLRYERDEGDKRNKKFELGEPNLPYYIDG 61
QY 98 DVKLTQSMALIRYIADKHNMLGSGPKERAEISMLEGAVLDIRGVSRIVASKDFETLKVD 157
DB 62 DVKLTQSMALIRYIADKHNMLGSGPKERAEISMLEGAVLDIRGVSRIVASKDFETLKVD 121
QY 158 FLSKLPMLKMFEDRLSHKTYTLNGDHYTHDPMLYDALDVLYLMDPCLDAPFKLYSFKK 217
DB 122 FLSKLPMLKMFEDRLCHKTYTLNGDHYTHDPMLYDALDVLYLMDPCLDAPFKLYCFKK 181
QY 218 RIRAIPOIDKYLLSKSYIAMPLOGWQATFGGSDHPKPSQSEALFQGGDGEPSQLPHCP 277
DB 182 RIRAIPOIDKYLLSKSYIAMPLOGWQATFGGSDHPKPS---DIVPRGSALRPSGAPG-P 237
QY 278 SVSPSAQPTWHPGOSQLF 295

```


CC According to the text of the specification, the C-terminal linker
 CC comprises 5 residues, however, only the sequence Aa Ser Ser END
 CC is given.
 CC The GST-derived protein domain allows rapid isolation using affinity
 CC chromatography to a solid phase contg. glutathione affixed to it.
 CC The Factor Xa specific cleavage site is provided when a subject
 CC fusion protein is produced in the vector pGEX-3X. The cleavage
 CC site is included to allow produ. of an NMBV structural protein
 CC free of the GST function domain.
 XX
 SQ Sequence 285 AA:
 Query Match 21.7%; Score 1164; DB 13; Length 285;
 Best Local Similarity 88.5%; Pred. No. 3e-100;
 Matches 224; Conservative 3; Mismatches 14; Indels 12; Gaps 3;
 QY 38 SPLIGYWKIKGLVOPTRLLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYIDG 97
 DB 2 SPLIGYWKIKGLVOPTRLLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYIDG 61
 QY 98 DVKLTQSMALIRYIADKHNNMGSPKERAISMLEGAVLDIRGVSRISAKDFETLKVD 157
 DB 62 DVKLTQSMALIRYIADKHNNMGSPKERAISMLEGAVLDIRGVSRISAKDFETLKVD 121
 QY 158 FLSKLPBEMKMFEDRLSHKTYLNGDHYTHDPFMYLDLVLYMDPCLDAFPKLYSFKK 217
 DB 122 FLSKLPBEMKMFEDRLCHKTYLNGDHYTHDPFMYLDLVLYMDPCLDAFPKLYCFKK 181
 QY 218 RIRAIPOIDKYLYKSKYIAMPLOGMOATFEGGDPHPKQSQOALFOGDDGSEPSQLPHCP 277
 DB 182 RIRAIPOIDKYLYKSKYIAMPLOGMOATFEGGDPHPKSD-----LIEG--RGIRN----- 229
 QY 278 SVSPSAQPTWHPG 290
 DB 230 SVSPSAQPTWHPG 242
 RESULT 12
 ID AAU10667 standard; Protein; 442 AA.
 AC AAU10667;
 DT 14-FEB-2002 (first entry)
 DE Human L1CAM Fn4-5 fusion polypeptide.
 KW Neurite outgrowth; fibronectin Type III repeat; cell adhesion molecule;
 KW Fn4-5; neurone; peripheral nerve damage; trauma; gliosis;
 KW Infectious; degenerative disease; malignant disease; antibacterial;
 KW central nervous system lesion; viruslike; antiparkinsonian; neurotropic;
 KW neuroprotective; antiinflammatory; human; L1CAM; mutant; mutein.
 OS Homo sapiens.
 OS Synthetic.
 PN US6313265-B1.
 PD 06-NOV-2001.
 PF 24-JUL-1995; 95US-0506296.
 PR 24-JUL-1995; 95US-0506296.
 PA (SCRI) SCRIPPS RES INST.
 PI Phillips G, Cunningham BA, Crossin KL;
 DR WPI; 2002-017011/02.
 PT Polypeptide for promoting neurite out-growth useful for treating
 PT diseases such as inflammation, Parkinson's disease, trauma, comprises
 PT fibronectin type III repeats derived from a family of cell adhesion

PT molecules -
 XX
 PS Claim 3; Fig 22; 132pp; English.
 XX
 CC The present invention relates to polypeptides that promote neurite
 CC growth. The polypeptides contain fibronectin Type III repeats derived
 CC from a family of cell adhesion molecules (CAMs). The polypeptides of the
 CC invention include the F80, 3-5, and 4-5 regions of the CAM family
 CC members chicken Ng-CAM, chicken Nr-CAM, mouse L1CAM and human L1CAM.
 CC The polypeptides of the invention are useful for promoting neurite
 CC outgrowth of neuronal cells in vitro e.g. in a cell culture system, or
 CC in vivo for treating disorders such as peripheral nerve damage
 CC associated with physical or surgical trauma, infection, bacterial or
 CC viral infections, toxin exposure, degenerative disease, malignant
 CC disease that affects peripheral or central neurones, or in surgical or
 CC transplantation methods in which new neuronal cells from brain, spinal
 CC cord or dorsal root ganglia are introduced and require stimulation of
 CC neurite outgrowth from the implant and innervation into the recipient
 CC tissue, where the diseases include central nervous systems lesions,
 CC gliosis, Parkinson's disease, Alzheimer's disease, gliotic response or
 CC inflammation. The present sequence represents human L1CAM Fn4-5 fusion
 CC polypeptide.
 XX
 SQ Sequence 442 AA:
 Query Match 21.7%; Score 1164; DB 23; Length 442;
 Best Local Similarity 75.2%; Pred. No. 6.2e-100;
 Matches 236; Conservative 7; Mismatches 33; Indels 38; Gaps 6;
 QY 38 SPLIGYWKIKGLVOPTRLLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYIDG 97
 DB 2 SPLIGYWKIKGLVOPTRLLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYIDG 61
 QY 98 DVKLTQSMALIRYIADKHNNMGSPKERAISMLEGAVLDIRGVSRISAKDFETLKVD 157
 DB 62 DVKLTQSMALIRYIADKHNNMGSPKERAISMLEGAVLDIRGVSRISAKDFETLKVD 121
 QY 158 FLSKLPBEMKMFEDRLSHKTYLNGDHYTHDPFMYLDLVLYMDPCLDAFPKLYSFKK 217
 DB 122 FLSKLPBEMKMFEDRLCHKTYLNGDHYTHDPFMYLDLVLYMDPCLDAFPKLYCFKK 181
 QY 218 RIRAIPOIDKYLYKSKYIAMPLOGMOATFEGGDPHPKQSQOALFOGDDGSEPSQLPHCP 276
 DB 182 RIRAIPOIDKYLYKSKYIAMPLOGMOATFEGGDPHPKSD-----LVPKSGHPALLEC 237
 QY 277 PSVSPSAQPT-----HP-----GOSQLFADLSREE-----LTVYMR 308
 DB 238 QSNVTSLLRQPPLSHNGVLTGYLSYHPDDEGKGQLSFLNLDPELRTNLTDLSPHLR 297
 QY 309 F-----LTQRLGPG 317
 DB 298 YRFQLQATTKEGPG 311
 RESULT 13
 ID AAY58917 standard; Protein; 1196 AA.
 AC AAY58917;
 DT 23-MAY-2000 (first entry)
 DE Fumonisin esterase-amine polyol amine oxidase fusion.
 KW Amino polyol amine oxidase; APAO; fumonisin esterase; mycotoxin;
 KW transgenic plant; detoxification; animal feed; silage;
 KW glutathione S transferase.
 XX
 OS Chimeric - Schistosoma japonicum.
 OS Chimeric - Bacteria.
 OS Chimeric - Exophiala splifera.
 PH key Location/Qualifiers

```

FT Peptide 1..229 /note="glutathione S transferase + polylinker"
FT /note- 230..1196
FT Protein /note="fimosin esterase-K:trAPAO fusion"
FT Misc-difference 734 /note="extra lysine"
FT Peptide 721..733 /note="peptide linker"
XX WO200004159-A1.
XX
XX 27-JAN-2000.
XX
XX 08-JUL-1999; 99MO-US15454.
XX
XX 15-JUL-1998; 98US-0092936.
XX
XX 21-MAY-1999; 99US-0135391.
XX
XX (PION-) PIONEER HI-BRED INT. INC.
XX (CURA-) CURAGEN CORP.
XX
XX Duvick JP, Gilliam JT, Maddox JR;
XX
XX MPI; 2000-182425/16.
XX
XX N-PSDB; AA258406.
XX
XX New isolated polynucleotides, polypeptides useful for detecting and
XX degrading fumonisin or structurally related mycotoxin in processed
XX grain or in silage
XX
XX Example 13; Page 129-131; 154pp; English.
XX
XX The present sequence is that of a fusion protein composed of
XX Schistosoma japonicum glutathione S transferase, and a bacterial
XX fumonisin esterase (BEST1) joined via a peptide linker to K:trAPAO
XX (see AA158909), a truncated, but functional, amino polyol amine
XX oxidase of E. spinifera. The construct was designed for expression
XX in Escherichia coli of a hybrid protein having BEST1 and APO
XX activity. The invention provides APO polynucleotides (see
XX AA258383-87) and polypeptides (see AA158900-05) of E. spinifera and
XX Rhinoclella atrovirens. The polynucleotides are used to
XX transform plant cells normally susceptible to Fusarium or other
XX toxin-producing fungus infection. Also provided are methods for
XX expressing APO, optionally as a fusion protein with fumosin
XX esterase, in transgenic plants, prokaryotic and non-plant
XX eukaryotic systems.) Methods for detoxification of grain, grain
XX processing, silage, food crops and in animal feed and rumen
XX microorganisms are also disclosed.
XX
XX Sequence 1196 AA:
XX
XX Query Match 21.7%; Score 1163; DB 21; Length 1196;
XX Best Local Similarity 38.9%; Pred. No. 3.9e-99;
XX Matches 325; Conservative 55; Mismatches 188; Indels 268; Gaps 31;
XX
XX 38 SPILGYKIKGLVOPTRLLLEYLEEKEEHLDERDECDKNNKFEELGLEPPNLPYITDG 97
XX |||||||
XX 2 SPILGYKIKGLVOPTRLLLEYLEEKEEHLDERDECDKNNKFEELGLEPPNLPYITDG 61
XX |||||||
XX 98 DVKLTGSMATIRYIADKNNMLGGSPKERAETISMEGAVLDIRGVSRIVASKDEFTLKV 157
XX |||||||
XX 62 DVKLTGSMATIRYIADKNNMLGGSPKERAETISMEGAVLDIRGVSRIVASKDEFTLKV 121
XX |||||||
XX 158 FLKSLPEMLKMFEDRLSHKRYLLNGDHTHDPMLYDALDVVLYDPMCLDAFPKLVSKK 217
XX |||||||
XX 122 FLKSLPEMLKMFEDRLSHKRYLLNGDHTHDPMLYDALDVVLYDPMCLDAFPKLVSKK 181
XX |||||||
XX 218 RIBAIPOIDIKYLSKSKYIAMPLOGMOTFGGDBPPKSSQSEALFQG----- 264
XX |||||||
XX 182 RIBAIPOIDIKYLSKSKYIAMPLOGMOTFGGDBPPKSSQSEALFQG----- 238
XX |||||||
XX 265 -----GDGGEPSQLPHCPVSP-----SAQPTWHPGOSQLFADLSREELTAVMR 308
XX |||||||

```

```

Db 239 LGQVGLAGDVMSFRGIPYAPPGVGLRMKPPQHARW-----AGVR 280
QY 309 FLVQRLGPGVLDAQAAP-----SDNCVFSVELQPPNA----- 342
Db 281 PATQ-FGSDCGAAYLKRKSLAPGVSEDCLY-LVWMA-PSGAKKPGQYVWVWVGSGFAGG 338
QY 343 -AALAHIDRGSPPPARALAVFFGRQPPRVSLVYGPLPHEPYMDVYVERHGFLPY 401
Db 339 TAAMPYYD--GEALARGVVWVTF-----NYRTNIIIGFFAHBG----- 374
QY 402 HRRPVLFQEYLDIDQMIENFRLPOASGLHHCCYKKHGRNLVMTTAPRGLOS----- 455
Db 375 -----LSRESPTGTS-----GNGLLDILALAKRWQSNARAFG 407
QY 456 GD--RAIWFGLYINISGAFGLAHVGLLELVNKAALDPARMTQKVFYQGRYDS----- 508
Db 408 GDPGRVTFVFGESAGASAIIGLLTS-----PLSKGLFRGAILLESPLTR 450
QY 509 -LAOLEAQFEAG-----LVNVVLIIDNGTGSWSLSKSPVPPGAPPLQFYP 553
Db 451 PLATLADSAASGERLADLSRLRSTDPATLMARADARPARSRDLRRRPPTGPIVDGHVLP 510
QY 554 QGPRFSVQGSRVASSLMTFSFGLGAFSGPRIFDVRFQGERLV-YEISLQ-----EALAIY 607
Db 511 QTDNAIATAAGQIAPV--RVLIGTNADBG-RAFLGRAPMETPADYQATLEAOFQQAAYA 567
QY 608 -----GNSPAAATTR-VYDGGFGWG-----KY-----TPPLTR 635
Db 568 ACYPLDGRATPKEMVARIFGDQNFNGVSAFSEALVRQGA-PVWRYXQNGNTEGGRAPATH 627
QY 636 GVDCPYLATYVDWHLFLESQAPKTRIRAPFCYFEQNO-GL-----PLRRHSDLYSH 685
Db 628 GAELIPV-----FGVFKLDELGLDWDPEEGPTPADRALGQLMSS 666
QY 686 YFGGLAETVLVVRMSSTLNLVYMDTVFHPGSAIEIRFATGYISSAFUGATGK 741
Db 667 AWVAFK-----NDPRAGDALTPMA-----YSTG--KSTMTFPGPCR 701

```

RESULT 14
AAV68853
ID AAV68853 standard; Protein; 1196 AA.
XX
AC AAV68853;
XX
DT 16-MAY-2000 (first entry)
XX
XX A fusion of aminopolylol amine oxidase/fumonisin esterase.
XX
KW Aminopolylol amine oxidase; APO; mycotoxin degradation; antimicrobial;
KW Plant fungal invasion; hydrogen peroxide; Fusarium; fungi;
XX fumonisin esterase.
XX
OS Synthetic.
OS Exophiala spinifera.
XX
FH Key Location/Qualifiers
FT Protein 1..229
FT Protein /note="GST and a linker" 230..721
FT Protein /note="bacterial fumonisin esterase" 722..733
FT Peptide /note="spacer sequence" 7340..1196
FT Protein /note="aminopolylol amine oxidase with an amino terminal lys for optimized expression"
XX
XX WO200004160-A1.
XX
XX 27-JAN-2000.
XX
XX 08-JUL-1999; 99MO-US15455.

```

PR 15-JUL-1998; 9805-0092936.
XX (PION-) PIONEER HI-BRED INT INC.
PA (CURA-) CURAGEN CORP.
XX
PI Crista OR, Duvick J, Folkerts O, Gilliam JT, Maddox JR;
DR WPI: 2000-182426/16.
XX N-PSDB; AAZ60645.
PT New nucleic acid encoding aminopolylol amine oxidase, used, e.g. to
XX generate plants resistant to Fusarium -
XX
PS Example 13; Page 124-126; 145pp; English.
XX
CC The present sequence represents a fusion protein of GST/Exophiala
CC splifera aminopolylol amine oxidase (APOO)/fumonisin esterase.
CC The APOO enzyme has homology to the flavin containing amine oxidase
CC family, that oxidise primary amine to an aldehyde or ketone, releasing
CC ammonia and hydrogen peroxide. The APOO enzyme degrades mycotoxins
CC that promote fungal invasion of plants. Destruction of mycotoxins by
CC APOO generates, as a by-product, hydrogen peroxide which is itself an
CC antimicrobial and stimulates the plants own defensive systems. The
CC APOO polynucleotides are used to generate plants (particularly maize)
CC that are resistant to Fusarium or other fungi that produce mycotoxins
CC and/or to degrade such mycotoxins (e.g. during ensiling); for
CC recombinant production of APOO polypeptides; as selection markers for
CC plant transformation; and to isolate related sequences from other
CC organisms. The APOO polypeptides are used to degrade mycotoxins in
CC plant materials, including expression in engineered bacteria and fungi,
CC e.g. rumen microflora.
XX
SQ Sequence 1196 AA;
XX
Query Match 21.7%; Score 1163; DB 21; Length 1196;
Best Local Similarity 38.9%; Pred. No. 3.9e-99;
Matches 325; Conservative 55; Mismatches 186; Indels 266; Gaps 31;
XX
QY 38 SPILGWKIKGLVQPTRLLEYLEEKEEHLERDEGDKWRNKKFELGEPNLPYYIDG 97
DB 2 SPILGWKIKGLVQPTRLLEYLEEKEEHLERDEGDKWRNKKFELGEPNLPYYIDG 61
XX
QY 98 DVKLTOSMAIIRYIADKHNLMGSPKERAISMLGCAVDIRGVSRIRYASKDEFELTKVD 157
DB 62 DVKLTOSMAIIRYIADKHNLMGSPKERAISMLGCAVDIRGVSRIRYASKDEFELTKVD 121
XX
QY 158 FLTKLPEMLKMFEDRLSHKTYLNGDHYTHPDEFLYDALDVLYLMDPCLDAFPKLYSKK 217
DB 122 FLTKLPEMLKMFEDRLCHNTYTLNGDHYTHPDEFLYDALDVLYLMDPCLDAFPKLYSKK 181
XX
QY 218 RIEAIPQIDIKYKSSKTYIMPLDQWQATFGGCDHPKSSQSEALFEG----- 264
DB 182 RIEAIPQIDIKYKSSKTYIMPLDQWQATFGGCDHPKSSQSEALFEG----- 238
XX
QY 265 -----GDGEPQLPHCPSPVSP-----SAOPWTHPQGSGQLFADLSREMLAYMR 308
DB 239 LQOVQGLADVMSFRCIPYAAAPRYVGLRKMKPPQAHARW-----AGVR 280
XX
QY 309 FLTORLPGILVDAQAARP-----SDNCVFSVELQLPKPA----- 342
DB 281 PNTQ-FGSDQCFGAAYLRKKSIAKGVSEDCILY-LNWAAPSGAKGQYRVWVWYGGGAGG 338
XX
QY 343 -AALAHLDNGSPPARREALIYFEGROPQPNVSELVVGPLPHPSYKNDVTVERHGGLPY 401
DB 339 TAAAPYD--GGAALAQGVVYTF-----NYRTNIIIGFAHAG----- 374
XX
QY 402 HRRPVLFQEVLDIDQMIFFNELPQASGLHHCFFYKHRGNLTMTTAPRGLOS----- 455
DB 375 -----LSRESPTGTS-----GNTGLDIIAALMWOSNARAFG 407
XX
QY 456 GD--RAATWFGLYNIGAGAFELHNVGLLELVNKKALDPAMTIOKVEYOGRYVDS----- 508
DB 408 GDGGRVTVGESAGASAIIGLLTS-----PLSKGLFRALIEESPOLTR 450

```

```

QY 509 -LAOLEAQEFAG-----LVNVVLIPDNGTGGSSLSKSPVPPGAPPIQIERYP 553
DB 451 PLATLADSAASGERLADLSRLSTDPATILMARADAPASRLRRPRTGPIVDGHVLP 510
XX
QY 554 QGPRFSVQGSRYASSLMTTSFGLGAFSPRIDVROGERLV-YEISLDQ-----EALYAI 607
DB 511 QDSAAIAAGALAPV--RVLIQTNADEG--RAFLGRAPMETPADYQAYLEAGFGDQAAVA 567
XX
QY 608 -----GGSNPAMATTR-YVDGFGMG-----KY-----TTPLTR 635
DB 568 ACYPRDGRATPKEMVARITGDQNFNGVAFSALVRQGAAPRVYQFNGTGGRAPATH 627
XX
QY 636 GVDCCPYLATYVDMHFLLESQAPRTIADAFCEVQNO-GL-----PLRRHSDLYSH 685
DB 628 GAELPYV-----FGVEFLDELGLFDWPEGPPTPADRALGOLMS 666
XX
QY 686 YFGGLAETVLVYRSMSTLNDYVMDTVFHPGAIIRRYANGYISSAFLEGATGK 741
DB 667 AWYRFKK-----NGDPAGDALTPA-----YSTG--KSTMTFGEGR 701
XX
RESULT 15
AA017115
ID AA017115 standard; Protein; 409 AA.
XX
AC AA017115;
XX
DT 11-JUN-2002 (first entry)
XX
DE GlutathioneS-transferase/osteoclast differentiation factor fusion protein.
XX
KW Mouse; cathepsin K promoter; bone metabolic disease; osteoporosis;
XX osteoclast differentiation factor; glutathione-S-transferase.
XX
OS Chimeric - Mus sp.
XX
OS Chimeric - Unidentified.
XX
PN JP2002034563-A.
XX
PD 05-FEB-2002.
XX
PE 19-JUL-2000; 2000JP-0218497.
XX
PR 19-JUL-2000; 2000JP-0218497.
XX
PA (SANY ) SANKYO CO LTD.
XX
DB WPI: 2002-275721/32.
XX
XX
XX Cathepsin K promoter useful for detecting a preventative agent for
XX treating bone metabolic diseases e.g., osteoporosis -
XX Disclosure; Page 14-15; 16pp; Japanese.
XX
CC The present invention provides the murine cathepsin K promoter. The
CC sequence can be used in the identification of treatments and preventive
CC agents for bone metabolic diseases such as osteoporosis. The present
CC sequence is a fusion protein of the murine osteoclast differentiation
CC factor and a glutathione-S-transferase described in the invention.
XX
SQ Sequence 409 AA;
XX
Query Match 21.7%; Score 1162.5; DB 23; Length 409;
Best Local Similarity 87.6%; Pred. No. 7.5e-100;
Matches 226; Conservative 1; Mismatches 24; Indels 7; Gaps 2;
XX
QY 38 SPILGWKIKGLVQPTRLLEYLEEKEEHLERDEGDKWRNKKFELGEPNLPYYIDG 97
DB 2 SPILGWKIKGLVQPTRLLEYLEEKEEHLERDEGDKWRNKKFELGEPNLPYYIDG 61
XX
QY 98 DVKLTOSMAIIRYIADKHNLMGSPKERAISMLGCAVDIRGVSRIRYASKDEFELTKVD 157

```

Db 62 DVKLTQSMATIRYIADKHNMIGGCPKERAIEISMLEGAVLDIRGVSRIRIAYSKOFETLKVD 121
QY 158 FLKLPMLKWFEDRLSHKTYLNGDHYTHPDPMLYDALDVVLYMDPMCLDAFPKLVSEFK 217
Db 122 FLKLPMLKWFEDRLCHKTYLNGDHYTHPDPMLYDALDVVLYMDPMCLDAFPKLVSEFK 181
QY 218 RIEAIPQIDRYLKSSKYIAMPLOGMOTFGGSDHPKKSQSEA-----LFGGDS-GEF 270
Db 182 RIEAIPQIDRYLKSSKYIAMPLOGMOTFGGSDHPKKSQSEA-----LFGGDS-GEF 241
QY 271 SQLPHCPSSVSPSAQPMTH 288
Db 242 SWLDVAQGRGKPEAOPFAH 259

Search completed: May 20, 2003, 11:34:02
Job time : 90.2615 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:30:54 ; Search time 32.1643 Seconds

(Without alignments)
912.939 Million cell updates/sec

Title: US-10-081-408-20

Perfect score: 5357
Sequence: 1 MDWLRLLEFLMAAQSINAA.....QAAACAPDLPFSGHGFSSHN 998

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 2942292 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents, AA: *
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6C.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/6D.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1180	22.0	447	US-08-506-296B-73	Sequence 73, Appl
2	1175.5	21.9	644	US-08-506-296B-65	Sequence 65, Appl
3	1175	21.9	536	US-08-974-180-20	Sequence 20, Appl
4	1171.5	21.9	514	US-08-974-349A-605	Sequence 605, Appl
5	1167.5	21.8	362	US-08-395-507-2	Sequence 2, Appl
6	1165	21.7	517	US-08-974-549A-606	Sequence 606, Appl
7	1164.5	21.7	442	US-08-506-296B-76	Sequence 76, Appl
8	1164	21.7	443	US-08-506-296B-70	Sequence 70, Appl
9	1163	21.7	1196	US-09-352-159-31	Sequence 31, Appl
10	1163	21.7	1196	US-09-352-168-31	Sequence 31, Appl
11	1162	21.7	272	US-08-910-820-4	Sequence 4, Appl
12	1162	21.7	397	US-08-327-874A-11	Sequence 11, Appl
13	1162	21.7	397	PCR-US94-09700-11	Sequence 11, Appl
14	1161	21.7	272	US-08-910-820-6	Sequence 6, Appl
15	1161	21.7	635	US-08-506-296B-71	Sequence 71, Appl
16	1158.5	21.6	426	US-08-737-248-4	Sequence 4, Appl
17	1158	21.6	439	US-08-506-296B-67	Sequence 67, Appl
18	1157.5	21.6	422	US-09-217-228-7	Sequence 7, Appl
19	1157.5	21.6	559	US-08-506-296B-66	Sequence 66, Appl
20	1157.5	21.6	647	US-08-305-764C-56	Sequence 56, Appl
21	1157	21.6	331	US-09-217-228-6	Sequence 6, Appl
22	1157	21.6	354	US-08-216-894-4	Sequence 4, Appl
23	1157	21.6	354	US-09-115-746-4	Sequence 4, Appl
24	1157	21.6	564	US-08-216-894-2	Sequence 2, Appl
25	1157	21.6	564	US-09-115-746-2	Sequence 2, Appl
26	1157	21.6	643	US-08-216-894-8	Sequence 8, Appl
27	1157	21.6	643	US-09-115-746-8	Sequence 8, Appl

28	1155.5	21.6	1140	US-08-471-112A-4	Sequence 4, Appl
29	1155	21.6	307	US-08-216-894-6	Sequence 6, Appl
30	1155	21.6	307	US-09-115-746-6	Sequence 6, Appl
31	1154.5	21.6	435	PCR-US95-04439-1	Sequence 1, Appl
32	1154	21.5	282	US-08-910-820-3	Sequence 3, Appl
33	1154	21.5	282	US-08-910-820-5	Sequence 5, Appl
34	1153	21.5	412	US-09-366-009-34	Sequence 34, Appl
35	1152.5	21.5	600	US-08-895-707-2	Sequence 2, Appl
36	1152.5	21.5	616	US-08-974-549A-602	Sequence 602, Appl
37	1152	21.5	632	US-08-506-296B-74	Sequence 74, Appl
38	1151.5	21.5	829	US-09-352-159-33	Sequence 33, Appl
39	1151.5	21.5	829	US-09-352-168-33	Sequence 33, Appl
40	1151	21.5	1205	US-09-352-159-29	Sequence 29, Appl
41	1151	21.5	1205	US-09-352-168-29	Sequence 29, Appl
42	1150.5	21.5	538	US-08-974-549A-602	Sequence 602, Appl
43	1150	21.5	472	US-08-216-894-10	Sequence 10, Appl
44	1150	21.5	472	US-09-115-746-10	Sequence 10, Appl
45	1149	21.4	542	US-08-506-296B-69	Sequence 69, Appl

ALIGNMENTS

RESULT 1
US-08-506-296B-73
; Sequence 73, Application US/08506296B
; Patent No. 6313265
; GENERAL INFORMATION:
; APPLICANT: Phillips, Greg
; APPLICANT: Cunningham, Bruce A.
; APPLICANT: Crossin, Kathryn L.
; TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES
; TITLE OF INVENTION: CONTAINING FIBRONECTIN TYPE III REPEATS AND METHODS OF USE
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESS: The Scripps Research Institute
; STREET: 10550 No. 6313265th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: California
; COUNTRY: U.S.
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/506,296B
; FILING DATE: 24-JUL-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: TSRI 488.0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 554-2937
; TELEFAX: (619) 554-6312
; INFORMATION FOR SEQ ID NO: 73:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
; US-08-506-296B-73
Query Match 22.0%; Score 1180; DB 4; Length 447;
Best Local Similarity 69.7%; Pred. No. 3.5e-107;
Matches 251; Conservative 9; Mismatches 50; Indels 50; Gaps 10;
DB 38 SPTIGWTKIKGLVOPRLLLEYLEEKYFEHLVERDGGDKWRNKKFELGLEFPPLPYIDG 97
2 SPTIGWTKIKGLVOPRLLLEYLEEKYFEHLVERDGGDKWRNKKFELGLEFPPLPYIDG 61

FILING DATE: 01-OCT-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Apple, Randolph Ted
 REGISTRATION NUMBER: 36,429
 REFERENCE/DOCKET NUMBER: 015389-002610US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 606:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 517 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FEATURE:
 NAME/KEY: Protein
 LOCATION: 1..517
 OTHER INFORMATION: /note="fusion protein composed of
 OTHER INFORMATION: glutathione-S-transferase sequence,
 OTHER INFORMATION: thrombin cleavage sequence, recognition
 OTHER INFORMATION: sequence for heart muscle protein
 OTHER INFORMATION: kinase, residues introduced by cloning
 OTHER INFORMATION: and HIRT protein fragment.
 US-08-974-549A-606

Query Match 21.7%; Score 1165; DB 4; Length 517;
 Best Local Similarity 66.9%; Pred. No. 1.3e-105;
 Matches 239; Conservative 14; Mismatches 48; Indels 56; Gaps 6;

QY 38 SPILGWIKIGLVQPTRLLEYLEEYERDEGDKWRNKKFELGEPNLPYIDG 97
 DB 2 SPILGWIKIGLVQPTRLLEYLEEYERDEGDKWRNKKFELGEPNLPYIDG 61
 QY 98 DVKLTOSMAIIRYIADKHNMLGSGPKERAISMLEGAVLDIRGVSRIRYSKPEFLKYD 157
 DB 62 DVKLTOSMAIIRYIADKHNMLGSGPKERAISMLEGAVLDIRGVSRIRYSKPEFLKYD 121
 QY 158 FLSKLEPMLEKMFEDRLSHKTYLNGDHVTHPDFMLYDALDVLYMDPCLDAPEKLVSPK 217
 DB 122 FLSKLEPMLEKMFEDRLSHKTYLNGDHVTHPDFMLYDALDVLYMDPCLDAPEKLVSPK 181
 QY 218 RIFAIPQIDIKYLLSKSYTIAMPLOGWQATFGGSDHPKSGSQEALFPGGSGEPPSOLPHC 277
 DB 182 RIFAIPQIDIKYLLSKSYTIAMPLOGWQATFGGSDHPKSGSQEALFPGGSGEPPSOLPHC 241
 QY 278 -----SVSPSAOPW-----PM-THPQ 291
 DB 242 RCRAVNSLSHYREVLPATFVRRLGPOGRLVQBGDPAFAFALVACLVCPWMDARPPA 301
 QY 292 SOLFADLS--REELTAVMRELTOR-----LQGLVDAQAARPSDNCVSEVELQ 339
 DB 302 APSEFQVSLCKELVAVLQRLCERGAKNVLAFFALDGAARGPPE-ATTSVASTYLP 357

RESULT 7
 US-08-506-296B-76
 Sequence 76, Application US/08506296B
 Patent No. 6313265
 GENERAL INFORMATION:
 APPLICANT: Phillips, Greg
 APPLICANT: Cunningham, Bruce A.
 APPLICANT: Crossin, Kathryn L.
 TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES
 TITLE OF INVENTION: CONTAINING FIBROBLAST TYPE III REPEATS AND METHODS OF USE
 NUMBER OF SEQUENCES: 77
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: The Scripps Research Institute
 STREET: 10550 No. 6313265th Torrey Pines Road, TPC-8
 CITY: La Jolla
 STATE: California
 COUNTRY: U.S.
 ZIP: 92037

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/506,296B
 FILING DATE: 24-JUL-1995
 CLASSIFICATION: 514
 ATTORNEY/AGENT INFORMATION:
 NAME: Fitting, Thomas
 REGISTRATION NUMBER: 34,163
 REFERENCE/DOCKET NUMBER: TSRI 488.0
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 554-2937
 TELEFAX: (619) 554-6312
 INFORMATION FOR SEQ ID NO: 76:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 443 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: Internal
 US-08-506-296B-76

Query Match 21.7%; Score 1164.5; DB 4; Length 443;
 Best Local Similarity 74.4%; Pred. No. 1.1e-105;
 Matches 236; Conservative 11; Mismatches 31; Indels 39; Gaps 7;

QY 38 SPILGWIKIGLVQPTRLLEYLEEYERDEGDKWRNKKFELGEPNLPYIDG 97
 DB 2 SPILGWIKIGLVQPTRLLEYLEEYERDEGDKWRNKKFELGEPNLPYIDG 61
 QY 98 DVKLTOSMAIIRYIADKHNMLGSGPKERAISMLEGAVLDIRGVSRIRYSKPEFLKYD 157
 DB 62 DVKLTOSMAIIRYIADKHNMLGSGPKERAISMLEGAVLDIRGVSRIRYSKPEFLKYD 121
 QY 158 FLSKLEPMLEKMFEDRLSHKTYLNGDHVTHPDFMLYDALDVLYMDPCLDAPEKLVSPK 217
 DB 122 FLSKLEPMLEKMFEDRLSHKTYLNGDHVTHPDFMLYDALDVLYMDPCLDAPEKLVSPK 181
 QY 218 RIFAIPQIDIKYLLSKSYTIAMPLOGWQATFGGSDHPKSGSQEALFPGGSGEPPSOLPHC 276
 DB 182 RIFAIPQIDIKYLLSKSYTIAMPLOGWQATFGGSDHPKSGSQEALFPGGSGEPPSOLPHC 237
 QY 277 PSVSPSAOPW-----THPGQSLFADLSREELTA-VMRELT-----TORLQGLVDAQAARPSD 328
 DB 238 PTLDSLTLEMGSPTR-----NGVLTSTYILAFQPLNTHELCP----- 275
 QY 329 NCYFVELQLPPKRAAL 345
 DB 276 ----LVEIRIPANESSL 288

RESULT 8
 US-08-506-296B-70
 Sequence 70, Application US/08506296B
 Patent No. 6313265
 GENERAL INFORMATION:
 APPLICANT: Phillips, Greg
 APPLICANT: Cunningham, Bruce A.
 APPLICANT: Crossin, Kathryn L.
 TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES
 TITLE OF INVENTION: CONTAINING FIBROBLAST TYPE III REPEATS AND METHODS OF USE
 NUMBER OF SEQUENCES: 77
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: The Scripps Research Institute
 STREET: 10550 No. 6313265th Torrey Pines Road, TPC-8
 CITY: La Jolla
 STATE: California
 COUNTRY: U.S.
 ZIP: 92037
 COMPUTER READABLE FORM:

```

1      MEDIUM TYPE: Floppy disk
2      COMPUTER: IBM PC compatible
3      OPERATING SYSTEM: PC-DOS/MS-DOS
4      SOFTWARE: PatentIn Release #1.0, Version #1.25
5      CURRENT APPLICATION DATA:
6      APPLICATION NUMBER: US/08/506,296B
7      FILING DATE: 24-JUL-1995
8      CLASSIFICATION: 514
9      ATTORNEY/AGENT INFORMATION:
10     NAME: Fitting, Thomas
11     REGISTRATION NUMBER: 34,163
12     REFERENCE/DOCKET NUMBER: TSRI 488.0
13     TELECOMMUNICATION INFORMATION:
14     TELEPHONE: (619) 554-2937
15     TELEFAX: (619) 554-6312
16     INFORMATION FOR SEQ ID NO: 70:
17     SEQUENCE CHARACTERISTICS:
18     LENGTH: 442 amino acids
19     TYPE: amino acid
20     TOPOLOGY: linear
21     MOLECULE TYPE: protein
22     FRAGMENT TYPE: internal
23
24     US-08-506-296B-70

```

Query Match	21.7%;	Score 1164;	DB 4;	Length 442;
Best Local Similarity	75.2%;	Pred. No. 1.3e-105;		
Matches 236; Conservative	7;	Mismatches 33;	Indels 38;	Gaps 6

QY	38	SPILGYNKIKLVOPTRLLEYLEEKXEEHLYEDEDQKNNKFFELGLEPNLPYIDG	97
Db	2	SPILGYNKIKLVOPTRLLEYLEEKXEEHLYEDEDQKNNKFFELGLEPNLPYIDG	61
QY	98	DVKLTQSMATIRYIADKHNMLGSSPKERAETISLEGAVLDIRGVSIYASKOETLKV	157
Db	62	DVKLTQSMATIRYIADKHNMLGSSPKERAETISLEGAVLDIRGVSIYASKOETLKV	129
QY	158	FLSKLPEMLKKMFEDRLSKHTYLYNDGDNHTNPDEFMLYDALDVLVYNDPSCDAFRLYEFKK	217
Db	122	FLSKLPEMLKKMFEDRLSKHTYLYNDGDNHTNPDEFMLYDALDVLVYNDPSCDAFRLYEFKK	184
QY	218	RLEALPOLDKTLKSSKTYIAMPLOGMOATFGGGDHPKPSQSOEALFGGGDGEPSQLP-NC	276
Db	182	RLEALPOLDKTLKSSKTYIAMPLOGMOATFGGGDHPKPSD---LVRGSPGHPREALLEEC	237
QY	277	PSVSPSAOPWT-----HP-----QOSOLFADLSREE-----LTAVMR	308
Db	238	QSNMSTLLRMQRP}SHNGVLTVGYLSYHPRDEGKGQGSFLRQRPDLRTNHLTDLSPHLR	297
QY	309	F-----LFGRLGPG 317	
Db	298	YRFQLOATTTKEGP 311	

```

: RESULT 9
: US-09-352-159-31
: Sequence 31, Application US/09352159A
: Patent No. 6211434
: GENERAL INFORMATION:
: APPLICANT: Duvick, Jonathan P.
: APPLICANT: Giliham, Jacob T.
: TITLE OF INVENTION: Amino Polyol Amine Oxidase
: FILE REFERENCE: Polynucleotides and Related Polypeptides and Methods of Use
: FILE REFERENCE: 1134
: CURRENT APPLICATION NUMBER: US/09/352,159A
: CURRENT FILING DATE: 1999-07-12
: EARLIER APPLICATION NUMBER: 60/092,936
: EARLIER FILING DATE: 1998-07-25
: EARLIER APPLICATION NUMBER: 60/135,391
: EARLIER FILING DATE: 1999-05-21
: NUMBER OF SEQ ID NOS: 46
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO: 31

```

```

; LENGTH: 1196
; TYPE: PRT
; ORGANISM: Unknown
US-09-352-159-31

```

Query Match	21.7%;	Score 1163;	DB 4;	Length 1196;
Best Local Similarity	38.9%;	Pred. No. 9.1e-105;		
Matches 325;	Conservative 55;	Mismatches 188;	Indels 268;	Gaps 31.

0Y	38	SPILGYMKIKLVOPTRLLEYLEEKEEENHLENEDECDKRNKKFELGFEFPMPLYIDG	97
0Y	38	SPILGYMKIKLVOPTRLLEYLEEKEEENHLENEDECDKRNKKFELGFEFPMPLYIDG	97
Db	2	SPILGYMKIKLVOPTRLLEYLEEKEEENHLENEDECDKRNKKFELGFEFPMPLYIDG	61
0Y	98	DVKLTQSAIIIRYIADKHNMLGSGPKERAISIMLEGAVIDIRGVSIAYSKDFELKYD	157
Db	62	DVKLTQSAIIIRYIADKHNMLGSGPKERAISIMLEGAVIDIRGVSIAYSKDFELKYD	121
0Y	158	FLSKLPKMLKMFEDRLSKITYLNDHDTYHPMLYDADLVLYNDPMLCDAPFKLYSFKK	217
Db	122	FLSKLPKMLKMFEDRLSKITYLNDHDTYHPMLYDADLVLYNDPMLCDAPFKLYSFKK	181
0Y	218	RIEAIPOIDKYLKSSKYTAMPLOQMOTFFGGDHPKPSQSOEALFQG-----	264
Db	182	RIEAIPOIDKYLKSSKYTAMPLOQMOTFFGGDHPKPS-----DLVPRSPPEFDOPVRRTD	238
0Y	265	-----GDGERSQLEPHCSVP-----SAQWTHPGOSQLEFADLSREBLAYMR	308
Db	239	LGQVGLAGLDVMSFSGRIYUAPAPVGLKMKRPPHARPW-----AGVR	280
0Y	309	FLTORLBSGLVDAQAR-----SDNCVFSVLOJRPKA-----	342
Db	281	PATQ-FGSDCFGAALIKRGSLAPVSEBCLY-LVNWAPSGKKPGQYVVMWYGGGFAAG	338
0Y	343	-AALAHLDRGSPPARRELAIVFFGROPOPVSELVYUPLPHPSYMRDIVERHGCLPY	401
Db	339	TAAAMYUD--GEALARQGVVYTF-----NYRNIIIGFRAHPC-----	374
0Y	402	HRRPVLFQOEYLDIDOMITNRELPOASGLLHHCCTCYKHRGNLYTMTPAPRGLOS-----	455
Db	375	-----LSRESPTGTS-----GNTGLDILALRNWOSNARAFG	407
0Y	456	GD--BATWFGLYUNISGAGFLHNHYGELLVNHKALDPARMTIOKVYOGGRYDS-----	508
Db	408	GDPRGYTFFGESAGASAIGLLTS-----PLKGLFRCALIESPOLTR	450
0Y	509	-LAOLEAOFEMG-----LVNVYLIPDNGTGGSWLSKSPVPPGAPPLQFYR	553
Db	451	PLATLADSAASGERLDADLSRLSTDPATMLAMARADAARPASRDRLRRPRTGPIVDGHVLP	510
0Y	554	QGPRESVQGSRYASSLMTFFSGCLAFSPRIFDVAFQGERLY-YEISIQ-----EALAIY	607
Db	511	QDSSAIAIAGQOLAP--RYLIGTNADEG-RAFLCRAPIETPADYQAVLEAOFQDAAVA	567
0Y	608	-----GGSNPAAMTTR-3YVGGGFGMG-----XY-----TTPLTR	635
Db	568	ACYPIDGRATPREMARIRFGDNOFIRGSAISALVROGARVYMRXYQFNGNTEGGRATH	627
0Y	636	GVDCEPLATVYDWHFLLESQAARTIRDAFCYFEOHQ-GL-----PLRRHHSIDLYSH	685
Db	628	GAIEIFYV-----FGVFKDELDELGLFMPPEEGPTPADRALAQGLMSS	666
0Y	686	YEGGLAEIVYVRSSTILNTDYWDVYFHHSGAIEIRFYATYGISSAFJGATGK	741
Db	667	AVWRBRK-----NGDPADALITWPA-----YSTG--KSTMTFGPEGR	701

RESULT 10
US-09-352-168-31
; Sequence 31, Application US/09352168A
; Patent No. 6211435
; GENERAL INFORMATION:
; APPLICANT: Crasta, Oswald R.
; APPLICANT: Duvick, Jonathan P.

APPLICANT: Folkerts, Otto
APPLICANT: Gilliam, Jacob T.
APPLICANT: Maddox, Joyce R.
TITLE OF INVENTION: Amino Polyl Amine Oxidase
TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
FILE REFERENCE: 08/75
CURRENT APPLICATION NUMBER: US/09/352,168A
CURRENT FILING DATE: 1999-07-12
EARLIER APPLICATION NUMBER: 60/092,936
EARLIER FILING DATE: 1998-07-25
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 31
LENGTH: 1196
TYPE: PRT
ORGANISM: Unknown
US-09-352-168-31

Query Match 21.7%; Score 1163; DB 4; Length 1196;
Best Local Similarity 38.9%; Pred. No. 9,1e-105;
Matches 325; Conservative 55; Mismatches 188; Indels 268; Gaps 31;

QY 38 SPILGWKIKGLVOPTRLLLEYLEKEEHLERDEGDKRNKKFELGLEFPNLPYIDG 97
DB 2 SPILGWKIKGLVOPTRLLLEYLEKEEHLERDEGDKRNKKFELGLEFPNLPYIDG 61
QY 98 DVKLTQSMATIRYIADKNNMLGSGPKERAEISMEGAVLDIRYGSRIAYSKDEFTLKYD 157
DB 62 DVKLTQSMATIRYIADKNNMLGSGPKERAEISMEGAVLDIRYGSRIAYSKDEFTLKYD 121
QY 158 FLKSLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVYLYNDPCLDAFPKLYCFKK 217
DB 122 FLKSLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVYLYNDPCLDAFPKLYCFKK 181
QY 218 RIEAIPQIDKYLKSSKIYAMPLOGMOATFGGDDHPKSGQSEALFQGG----- 264
DB 182 RIEAIPQIDKYLKSSKIYAMPLOGMOATFGGDDHPKSG--DLVPSKSPETTDPPVRRTD 238
QY 265 -----GDGGEPSQLPHCPSPV-----SAQPMTHGQSOLFADLSREELTAVMR 308
DB 239 LGQVQGLAGDVMSFRGIPYAAPVYGLRWKPPQIARPM-----AGVR 280
QY 309 FLTORLCPGLVDAQAARP-----SDNCVFSVELQLPRA----- 342
DB 281 PATD-FGSDCFGAAYLRKKSGLAPGVSEDCLY-LNVMAFSGAKPCQPYVMVWVYGGPAGG 338
QY 343 -AALAHLDGRSPPARREALIVFGRQPPRVSELVYGLPSPRYMDVTERHGGPLPY 401
DB 339 TAAMPYTD--GEALARGVVVYTF-----NYRTNIIIGFFAHPC----- 374
QY 402 HRRFVLFOEYLDIDQMTFNRFLPOASGLLHHCFCYKHRGNLVMTTAPRGLQS----- 455
DB 375 -----LSRESPTGTS-----GNYGLDLILALRVVQSNARAFG 407
QY 456 GD--RAIFWGLIYINISGAGFELHNVGLELLVNHKALDPARTYIKVYQGRYDS----- 508
DB 408 GDRPRVYVFGESAGASAIIGLLTS-----PLSKGFLRGAILSPGLTR 450
QY 509 -LAQLENOFEG-----LVNVYLIIPDNGTGSMSLSKSPVPPAPRPLQYTP 553
DB 451 PLATLADSAASGERLADLSRLRSTDPATTLARADAPARSDLRPRPTPIYDGHVLP 510
QY 554 QCPFFSVGSRVASSLTFSEGLAFSGPRIFDVFOGERLV-YEISIQ-----EALAIY 607
DB 511 QTSBAIAIAGGLAV--RVILGTINADEG-RAFLGARPMETRADYQATYLEAQFGQAALVA 567
QY 608 -----GNSPAAMTTR-YVDGCFQMG-----KY-----TTPLTR 635
DB 568 ACYPLDGRATPEKEMARIFGDNQFNRGVSASEALVROGAPVWYRQFNGNTEGGRAPATH 627
QY 636 GVDCPYLATYVDHMFLESQAPKTRDAFCVFEQNO-GL-----PLRRHSDLXSH 685
DB 628 GAELIPYV-----FGVFKLDELGLFDMPPREGPTPADRALGQLMSS 666

QY 686 YFGCLAETLYVRSMSSTLNYDVTVEHPHSCAIEIRFYATGYISSAFLEGATGK 741
DB 667 AMVRFK-----NGDPAGDALTWPA-----YSTG--KSTMTFGEGGR 701

RESULT 11
US-08-910-820-4
Sequence 4, Application US/08910820
Patent No. 6258579
GENERAL INFORMATION:

APPLICANT: Mercurio, Frank
APPLICANT: Zhu, Hengyi
APPLICANT: Barbosa, Miguel
APPLICANT: Li, Gian
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
TITLE OF INVENTION: COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Makl, David J. 31,392
REGISTRATION NUMBER: 860098, 41361
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-910-820-4

Query Match 21.7%; Score 1162; DB 4; Length 272;
Best Local Similarity 80.8%; Pred. No. 8.6e-106;
Matches 227; Conservative 4; Mismatches 32; Indels 18; Gaps 2;

QY 38 SPILGWKIKGLVOPTRLLLEYLEKEEHLERDEGDKRNKKFELGLEFPNLPYIDG 97
DB 2 SPILGWKIKGLVOPTRLLLEYLEKEEHLERDEGDKRNKKFELGLEFPNLPYIDG 61
QY 98 DVKLTQSMATIRYIADKNNMLGSGPKERAEISMEGAVLDIRYGSRIAYSKDEFTLKYD 157
DB 62 DVKLTQSMATIRYIADKNNMLGSGPKERAEISMEGAVLDIRYGSRIAYSKDEFTLKYD 121
QY 158 FLKSLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVYLYNDPCLDAFPKLYCFKK 217
DB 122 FLKSLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVYLYNDPCLDAFPKLYCFKK 181
QY 218 RIEAIPQIDKYLKSSKIYAMPLOGMOATFGGDDHPKSGQSEALFQGGDGEPSQLPHCP 277
DB 182 RIEAIPQIDKYLKSSKIYAMPLOGMOATFGGDDHPKSDPREFITDMAG-----VACL 235
QY 278 SVSPSAOPWTHGQSOLFADLSREELTAVMRFLVQRLPGL 318
DB 236 KGTADADEWCDSGLSLGPDAAAPG-----GPGL 264

RESULT 12
US-08-327-874A-11
Sequence 11, Application US/08327874A
Patent No. 6372249
GENERAL INFORMATION:
APPLICANT: BAYLOR COLLEGE OF MEDICINE
APPLICANT: SMITH, JAMES R.
APPLICANT: DRUTZ, DAVID J.
APPLICANT: WILSON, DEBORAH R.
APPLICANT: ZUMSTEIN, LOUIS A.
TITLE OF INVENTION: SENESCENT CELL DERIVED INHIBITORS OF
TITLE OF INVENTION: DNA SYNTHESIS
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROGERS & WELLS
STREET: 200 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10166
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/327,874A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/808,523
FILING DATE: 16-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/970,462
FILING DATE: 02-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/113,372
FILING DATE: 30-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,564
FILING DATE: 17-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/203,535
FILING DATE: 25-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/229,420
FILING DATE: 15-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/274,535
FILING DATE: 13-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT US94/09700
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/327,874
FILING DATE: 24-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: NORTON, GERARD P.
REGISTRATION NUMBER: 36,621
REFERENCE/DOCKET NUMBER: 3634-8-CIP10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 878-3148
TELEFAX: (212) 878-8375
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 397 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
IMMEDIATE SOURCE:

CLONE: GST-SDI-1 fusion protein
US-08-327-874A-11
Query Match 21.7%; Score 1162; DB 4; Length 397;
Best Local Similarity 83.9%; Pred. No. 1.7e-105;
Matches 224; Conservative 10; Mismatches 29; Indels 4; Gaps 2;
QY 38 SPILGYKIRKIGLVOPRTLLLEYLEEKYEELHYERDECDKRNKKFELGLFPNLPYIDG 97
DB 2 SPILGYKIRKIGLVOPRTLLLEYLEEKYEELHYERDECDKRNKKFELGLFPNLPYIDG 61
QY 98 DVKLTOGSMALIRYADKNNMIGSPKERAETISMEGAVLDIRYGVSRIRAYSKDEFELKVD 157
DB 62 DVKLTOGSMALIRYADKNNMIGSPKERAETISMEGAVLDIRYGVSRIRAYSKDEFELKVD 121
QY 158 FLTKLPMLKMFEDRLSHKTYLNGDHTYHPDPMILYDALDVLYLMDPKLAFPKLYSFKK 217
DB 122 FLTKLPMLKMFEDRLCHKTYLNGDHTYHPDPMILYDALDVLYLMDPKLAFPKLYSFKK 181
QY 218 RIEAIPQIDKYLKSSKYIAMPLOGMQATFGGDDHPPKSSQSEALFQGGDGEPSQLPHCP 277
DB 182 RIEAIPQIDKYLKSSKYIAMPLOGMQATFGGDDHPPKSSQSEALFQGGDGEPSQLPHCP 277
QY 278 SVSPSAQPMWTHPGQSOLFADLSREELT 304
DB 238 AGDVQNPCCSKACRRLFGPVDSEQLS 264
RESULT 13
PCT-US94-09700-11
Sequence 11, Application PC/TUS9409700
GENERAL INFORMATION:
APPLICANT: BAYLOR COLLEGE OF MEDICINE
APPLICANT: SMITH, JAMES R.
TITLE OF INVENTION: SENESCENT CELL DERIVED INHIBITORS OF
TITLE OF INVENTION: DNA SYNTHESIS
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOWREY & SIMON
STREET: 1299 PENNSYLVANIA AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09700
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/808,523
FILING DATE: 16-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/970,462
FILING DATE: 02-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/113,372
FILING DATE: 30-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,564
FILING DATE: 17-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/203,535
FILING DATE: 25-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/229,420
FILING DATE: 15-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/274,535

FILED DATE: 13-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: AUERBACH, JEFFREY I.
REGISTRATION NUMBER: 32,680
REFERENCE/DOCKET NUMBER: 225-102-C1P7-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 383-7451
TELEFAX: (202) 383-6610
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 397 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
IMMEDIATE SOURCE:
CLONE: GST-SDI-1 fusion protein
PCT-US94-09700-11

Query Match 21.7%; Score 1162; DB 5; Length 397;
Best Local Similarity 83.9%; Pred. No. 1.7e-105;
Matches 224; Conservative 10; Mismatches 29; Indels 4; Gaps 2;

QY 38 SPLIGTWKIKGLVQPTRLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYTI DG 97
DB 2 SPLIGTWKIKGLVQPTRLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYTI DG 61
QY 98 DVKLTOSMAIRIYADKHNMLGSGPKERAEISMLEGAVLDIRYGSRIATSKDFETLKVD 157
DB 62 DVKLTOSMAIRIYADKHNMLGSGPKERAEISMLEGAVLDIRYGSRIATSKDFETLKVD 121
QY 158 FLKSLPEMLKMFEDRLSHKTYLNGDHVTHDPFMYDALDVLYMDPCLDAPFKLYSFKK 217
DB 122 FLKSLPEMLKMFEDRLCHKTYLNGDHVTHDPFMYDALDVLYMDPCLDAPFKLYSFKK 181
QY 218 RIRAIPIIDKYLKSSKIAPLOGWMTFGGSDHPKPSQSQEALFOGDDGGEPSQLPHCP 277
DB 182 RIRAIPIIDKYLKSSKIAPLOGWMTFGGSDHPKPSQSQEALFOGDDGGEPSQLPHCP 237
QY 278 SVSPSAQPTWTHPGSQLEFADLSREELT 304
DB 238 AGDYRQNPCCSKACRRIRFGPVDSQSL 264

RESULT 14
US-08-910-820-6
Sequence 6, Application US/08910820
Patent No. 6258579
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
APPLICANT: Zhu, Hengyi
APPLICANT: Barbosa, Miguel
APPLICANT: Li, Gian
APPLICANT: Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
TITLE OF INVENTION: COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: MAKI, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-910-820-6

Query Match 21.7%; Score 1161; DB 4; Length 272;
Best Local Similarity 80.8%; Pred. No. 1.1e-105;
Matches 227; Conservative 4; Mismatches 32; Indels 18; Gaps 2;

QY 38 SPLIGTWKIKGLVQPTRLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYTI DG 97
DB 2 SPLIGTWKIKGLVQPTRLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYTI DG 61
QY 98 DVKLTOSMAIRIYADKHNMLGSGPKERAEISMLEGAVLDIRYGSRIATSKDFETLKVD 157
DB 62 DVKLTOSMAIRIYADKHNMLGSGPKERAEISMLEGAVLDIRYGSRIATSKDFETLKVD 121
QY 158 FLKSLPEMLKMFEDRLSHKTYLNGDHVTHDPFMYDALDVLYMDPCLDAPFKLYSFKK 217
DB 122 FLKSLPEMLKMFEDRLCHKTYLNGDHVTHDPFMYDALDVLYMDPCLDAPFKLYSFKK 181
QY 218 RIRAIPIIDKYLKSSKIAPLOGWMTFGGSDHPKPSQSQEALFOGDDGGEPSQLPHCP 277
DB 182 RIRAIPIIDKYLKSSKIAPLOGWMTFGGSDHPKPSQSQEALFOGDDGGEPSQLPHCP 235
QY 278 SVSPSAQPTWTHPGSQLEFADLSREELTAVRFLTORLPGSL 318
DB 236 GRTADDEKCDAGLGPDAAG-----GPGSL 264

RESULT 15
US-08-506-296B-71
Sequence 71, Application US/08506296B
Patent No. 6313265
GENERAL INFORMATION:
APPLICANT: Phillips, Greg
APPLICANT: Cunningham, Bruce A.
APPLICANT: Crossin, Kathryn L.
TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES
TITLE OF INVENTION: CONTAINING FIBRONECTIN TYPE III REPEATS AND METHODS OF USE
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute
STREET: 10550 No. 6313265th Torrey Pines Road, TPC-8
City: La Jolla
STATE: California
COUNTRY: U.S.
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/506,296B
FILING DATE: 24-JUL-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: TSRI 488.0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 554-2937

TELEFAX: (619) 554-6312
INFORMATION FOR SEQ ID NO: 71:
SEQUENCE CHARACTERISTICS:
LENGTH: 635 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: C-terminal
US-08-506-296B-71

Query Match 21.7%; Score 161; DB 4; Length 635;
Best Local Similarity 61.6%; Pred. No. 4.7e-105;
Matches 252; Conservative 11; Mismatches 51; Indels 95; Gaps 11;

QY 38 SPILGWKIKGLVOPRLLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYIDG 97
DB 2 SPILGWKIKGLVOPRLLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYIDG 61
QY 98 DVKLTQSMATIRIADKHNMLGSGPKERAEISMLEGAVLDIRYGVSRIRAYSKDETLKVD 157
DB 62 DVKLTQSMATIRIADKHNMLGSGPKERAEISMLEGAVLDIRYGVSRIRAYSKDETLKVD 121
QY 158 FLSKLPBMLKMFEDRLSHKTYLNGDHVTHPDFMLYDALDVLYMDPMLDAFPRLVSFKK 217
DB 122 FLSKLPBMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVLYMDPMLDAFPRLVCFKK 181
QY 218 RIRAIPOIDKYLKSKYIAMPLOGMQATFEGGDHPFKS-----OSQE 259
DB 182 RIRAIPOIDKYLKSKYIAMPLOGMQATFEGGDHPFKSDLVPRGSHIHKSHIVVPANTTS 241
QY 260 ALFQG-----GDGGEPSQLP-HCPSPSPSAQPWT 287
DB 242 AILSGLRPYSSYHVEVOAFNGRGLGRASEWTFSTPEGVPGHPEALHLECOSDTSLLHMQ 301
QY 288 -----HP--GQS--QLFADLSREELTAVMRFLTQRLGPGLVDAQAQAPS 327
DB 302 PPLSHNGVLTYGLSYHPVEGESKEOLFENLSDPELRT--HNLV-NLNPDL----- 349
QY 328 DNCVFSEVLQPPKAALAHLDGSPPPA--REALAVFEGROPQPNVS 374
DB 350 ---QYRFQLO-----ATTQGGPQQAIVREGGTMALFGKPDPEGNIS 387

Search completed: May 20, 2003, 11:41:03
Job time : 37.1643 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:37:59 ; Search time 159.01 Seconds

(without alignments)
622.443 Million cell updates/sec

Title: US-10-081-408-20

Perfect score: 5357

Sequence: 1 MDWLRLFLFMAAOSINAA.....QAACAPDLPAFSGGFSHN 998

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5357	100.0	998	9	US-10-081-408-20
2	3667	74.1	763	9	US-10-081-408-2
3	2480.5	46.3	729	10	US-09-919-497-51
4	1165.5	21.8	401	9	US-10-190-866A-1
5	1162	21.7	272	9	US-09-844-988-4
6	1162	21.7	272	10	US-09-844-908-4
7	1161	21.7	272	9	US-09-844-988-6
8	1161	21.7	272	10	US-09-844-908-6
9	1160.5	21.7	298	9	US-09-844-908-6
10	1159	21.6	218	9	US-09-910-600-26
11	1157.5	21.6	422	9	US-09-824-438-7
12	1157	21.6	331	9	US-09-824-438-6
13	1155.5	21.6	1140	9	US-09-950-634-4
14	1154	21.5	282	9	US-09-844-988-3
15	1154	21.5	282	9	US-09-844-988-5
16	1154	21.5	282	10	US-09-844-908-3
17	1154	21.5	282	10	US-09-844-908-5
18	1153	21.5	412	9	US-09-775-964-34
19	1152	21.5	364	9	US-09-788-268-14

20	1151	21.5	239	10	US-09-823-153-7	Sequence 7, Appl1
21	1150	21.5	245	9	US-10-267-311-23	Sequence 23, Appl1
22	1150	21.5	394	10	US-09-990-578-4	Sequence 4, Appl1
23	1148.5	21.4	354	10	US-09-823-153-8	Sequence 8, Appl1
24	1148	21.4	324	9	US-10-267-311-25	Sequence 25, Appl1
25	1148	21.4	348	9	US-09-810-600-22	Sequence 22, Appl1
26	1148	21.4	348	9	US-09-910-600-23	Sequence 23, Appl1
27	1148	21.4	348	9	US-09-910-600-24	Sequence 24, Appl1
28	1148	21.4	348	9	US-09-910-600-25	Sequence 25, Appl1
29	1145	21.4	649	9	US-10-174-784-9	Sequence 9, Appl1
30	1144	21.4	218	9	US-10-081-408-4	Sequence 4, Appl1
31	1482.5	9.0	229	9	US-10-102-806-554	Sequence 554, App
32	196.5	3.7	222	9	US-09-784-739-4	Sequence 4, Appl1
33	196.5	3.7	220	10	US-09-925-301-1318	Sequence 1318, Ap
34	179.5	3.4	222	9	US-09-784-739-3	Sequence 3, Appl1
35	173.5	3.2	222	9	US-09-784-739-5	Sequence 5, Appl1
36	169.5	3.2	203	9	US-09-847-208-48	Sequence 48, Appl1
37	164.5	3.1	222	9	US-09-784-739-1	Sequence 1, Appl1
38	109.5	2.0	240	9	US-09-864-899-23	Sequence 23, Appl1
39	106	2.0	233	10	US-09-765-213A-2	Sequence 2, Appl1
40	106	2.0	617	9	US-10-005-530-58	Sequence 58, Appl1
41	104	1.9	980	9	US-09-908-193-22	Sequence 22, Appl1
42	103	1.9	2224	9	US-10-115-563-14	Sequence 14, Appl1
43	102	1.9	1194	9	US-09-738-826-5832	Sequence 5832, Ap
44	101	1.9	617	9	US-10-005-530-46	Sequence 46, Appl1
45	99	1.8	233	10	US-09-765-213A-4	Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-10-081-408-20
; Sequence 20, Application US/10081408
; Patent No. US20020160482A1
; GENERAL INFORMATION:
; APPLICANT: Abtams, n, Lars
; TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
; FILE REFERENCE: 13425-053001
; CURRENT APPLICATION NUMBER: US/10/081.408
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: SE 010625-3
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 60/272,247
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 998
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant construct
US-10-081-408-20

Query Match 100.0%; Score 5357; DB 9; Length 998;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 998; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDWLRLFLFMAAOSINAAOHDAVNRKFKEDQNASPIIGYKIKGLVOPFLLEYL 60
DB 1 MDWLRLFLFMAAOSINAAOHDAVNRKFKEDQNASPIIGYKIKGLVOPFLLEYL 60
QY EEKKEEHLVYRDEDDKRNKKFELGEPNLPYIIDGVKLTOSMATIRYADHNHMG 120
DB EEKKEEHLVYRDEDDKRNKKFELGEPNLPYIIDGVKLTOSMATIRYADHNHMG 120
QY 121 SPKERABISMLEGAVLDIRGVSRIRAYSKDFELTKVDFLSKLPMLKMFEDRLSHKTYLN 180
DB 121 SPKERABISMLEGAVLDIRGVSRIRAYSKDFELTKVDFLSKLPMLKMFEDRLSHKTYLN 180
QY 181 GDHYTHDFMLYDALDVYLLTMDPMLCLDAFRLVSKFKRIEAIPOIDIXTKLSSKTIAMPLO 240

```

Db 181 GDHVTHTDFHLYDLVDLVVLMDDPKCLAPFKLVSEFKRIEALPDIDKYLKSSKTIAPLQ 240
QY 241 GWOATFGGCHPRKPSQSQEALFQGGDGEPSQLPHCPVSFSAQPTWHPGQSOLFADLSR 300
Db 241 GWOATFGGCHPRKPSQSQEALFQGGDGEPSQLPHCPVSFSAQPTWHPGQSOLFADLSR 300
QY 301 EELTAVNRFLTORLGPGLVYAAOARPSDNCVFSEYELDPPRAALAHLDKRGSPAREAL 360
Db 301 EELTAVNRFLTORLGPGLVYAAOARPSDNCVFSEYELDPPRAALAHLDKRGSPAREAL 360
QY 361 AIVFEGROPVNSSELYVGLPHPSYMRDVTVERHGGPLPYHRRPVLFOEYLDIDOMIFN 420
Db 361 AIVFEGROPVNSSELYVGLPHPSYMRDVTVERHGGPLPYHRRPVLFOEYLDIDOMIFN 420
QY 421 RELPQASGLLHHCFFYKRRGNLYTNTTAPRGLOSGBRATFGLYINISGAGFLHNGL 480
Db 421 RELPQASGLLHHCFFYKRRGNLYTNTTAPRGLOSGBRATFGLYINISGAGFLHNGL 480
QY 481 EELVNHKALDPAKWTIOKVFQGRYDSLAOLEAOFEGALVNVVLPDNGTGSWSLKS 540
Db 481 EELVNHKALDPAKWTIOKVFQGRYDSLAOLEAOFEGALVNVVLPDNGTGSWSLKS 540
QY 541 VPPGAPPLQFYPOGPRFVSQGSFVASSLMTFSGLGAFSGPRIFDVRFQGERLYEISL 600
Db 541 VPPGAPPLQFYPOGPRFVSQGSFVASSLMTFSGLGAFSGPRIFDVRFQGERLYEISL 600
QY 601 QEALATVIGNSPAMATRYVDGFGMGKTYTPTLRGVDCPLATYVDVHFLLESQAPRTI 660
Db 601 QEALATVIGNSPAMATRYVDGFGMGKTYTPTLRGVDCPLATYVDVHFLLESQAPRTI 660
QY 661 RDAFCVEQONGPLRLRHSDLYSHYFGLAEFVLYVRSMTLLNDYVMDPTVFHPSGAI 720
Db 661 RDAFCVEQONGPLRLRHSDLYSHYFGLAEFVLYVRSMTLLNDYVMDPTVFHPSGAI 720
QY 721 EIRFYATGYISSAFLEFGATGKYGNVSEHTLGTVHTSHAFKVDVDAAGLENNVVAEDMY 780
Db 721 EIRFYATGYISSAFLEFGATGKYGNVSEHTLGTVHTSHAFKVDVDAAGLENNVVAEDMY 780
QY 781 FVPAVWPSPHOLQRLQVTRKLEMEQOAAFVGSATPRLYTLASNSNMKGHPRGYRI 840
Db 781 FVPAVWPSPHOLQRLQVTRKLEMEQOAAFVGSATPRLYTLASNSNMKGHPRGYRI 840
QY 841 QMLSFAGEPLPONSMMARGFSEMERQOLAVTQRKEEPSSSSVFNQNDMPATVDFSDFIN 900
Db 841 QMLSFAGEPLPONSMMARGFSEMERQOLAVTQRKEEPSSSSVFNQNDMPATVDFSDFIN 900
QY 901 NETLAGKDLVAMVTAGFLHHPAEDIPNTVYVGNVGFLLRPYNFDEDPSTYSADSIYF 960
Db 901 NETLAGKDLVAMVTAGFLHHPAEDIPNTVYVGNVGFLLRPYNFDEDPSTYSADSIYF 960
QY 961 RGDODAGACEVNPPLACLPOAAACAPDLPAFSGGFSHN 998
Db 961 RGDODAGACEVNPPLACLPOAAACAPDLPAFSGGFSHN 998

RESULT 2
US-10-081-408-2
; Sequence 2, Application US/10081408
; Patent No. US20020160482A1
; GENERAL INFORMATION:
; APPLICANT: Abrahams, n, Lars
; APPLICANT: Nilsson, Joakim
; TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
; FILE REFERENCE: 13425-053001
; CURRENT APPLICATION NUMBER: US/10/081.408
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: SE 0100625-3
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 60/272.247
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
```

```

; SEQ ID NO 2
; LENGTH: 763
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-081-408-2

Query Match          74.1%; Score 3967; DB 9; Length 763;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 735; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 263 QGGDGEPSQLPHCPVSFSAQPTWHPGQSOLFADLSREELTAVNRFLTORLGPGLVDA 322
Db 263 RGGDGEPSQLPHCPVSFSAQPTWHPGQSOLFADLSREELTAVNRFLTORLGPGLVDA 322
QY 323 QARSDNCVFSEYELDPPRAALAHLDKRGSPAREALATVYFGRGROPVNSSELYVGLP 382
Db 323 QARSDNCVFSEYELDPPRAALAHLDKRGSPAREALATVYFGRGROPVNSSELYVGLP 382
QY 383 HPSYMRDVTVERHGGPLPYHRRPVLFOEYLDIDOMIFNRELPOASGLLHHCFFYKRRGN 442
Db 383 HPSYMRDVTVERHGGPLPYHRRPVLFOEYLDIDOMIFNRELPOASGLLHHCFFYKRRGN 442
QY 443 LVTMTTAPRGLOSGBRATFGLYINISGAGFLHNGVLELLVNHKALDPAKWTIOKVFYQ 502
Db 443 LVTMTTAPRGLOSGBRATFGLYINISGAGFLHNGVLELLVNHKALDPAKWTIOKVFYQ 502
QY 503 GRYVDSLQLEAOFEGALVNVVLPDNGTGSWSLKSVPVPPAPPLQFYPOGPRFVSQ 562
Db 503 GRYVDSLQLEAOFEGALVNVVLPDNGTGSWSLKSVPVPPAPPLQFYPOGPRFVSQ 562
QY 563 SRVASSLMTFSGLGAFSGPRIFDVRFQGERLYEISLQEALATVIGNSPAMATRYVDG 622
Db 563 SRVASSLMTFSGLGAFSGPRIFDVRFQGERLYEISLQEALATVIGNSPAMATRYVDG 622
QY 623 GFGMKYTTPTLRGVDCPYLATYVDVHFLLESQAPKTRDAFCVEQONGPLRLRHSDL 682
Db 623 GFGMKYTTPTLRGVDCPYLATYVDVHFLLESQAPKTRDAFCVEQONGPLRLRHSDL 682
QY 683 YSHYFGLAEFVLYVRSMTLLNDYVMDPTVFHPSGAIERFYATGYISSAFLEFGATGY 742
Db 683 YSHYFGLAEFVLYVRSMTLLNDYVMDPTVFHPSGAIERFYATGYISSAFLEFGATGY 742
QY 743 GNOVSEHTLGTVHTSHAFKVDVDAAGLENNVVAEDMYFVPAVWPSPHOLQRLQVTRK 802
Db 743 GNOVSEHTLGTVHTSHAFKVDVDAAGLENNVVAEDMYFVPAVWPSPHOLQRLQVTRK 802
QY 803 LLEMEQOAAFVGSATPRLYTLASNSNMKGHPRGYRIOMLSFAGEPLPONSMMARGFSW 862
Db 803 LLEMEQOAAFVGSATPRLYTLASNSNMKGHPRGYRIOMLSFAGEPLPONSMMARGFSW 862
QY 863 ERYQOLAVTQRKEEPSSSSVFNQNDMPATVDFSDFINNETTAGKDLVAMVTAGFLHHP 922
Db 863 ERYQOLAVTQRKEEPSSSSVFNQNDMPATVDFSDFINNETTAGKDLVAMVTAGFLHHP 922
QY 923 AEDIPNTVYVGNVGFLLRPYNFDEDPSTYSADSIYFRGDODAGACEVNPPLACLPOAAA 982
Db 923 AEDIPNTVYVGNVGFLLRPYNFDEDPSTYSADSIYFRGDODAGACEVNPPLACLPOAAA 982
QY 983 CAPDLPAFSGGFSHN 998
Db 983 CAPDLPAFSGGFSHN 998

RESULT 3
US-09-919-497-51
; Sequence 51, Application US/09919497
; Patent No. US20020106662A1
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919.497
; CURRENT FILING DATE: 2001-07-31
```

; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 51
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-919-497-51

Query Match 46.3%; Score 2480.5; DB 10; Length 729;
Best Local Similarity 64.1%; Pred. No. 2.4e-203;
Matches 468; Conservative 86; Mismatches 145; Indels 31; Gaps 3;

OY 268 GEPQLPHCPSPSPAPOTPHGQSLFADLSREELTAVMRFOTPLGGLVDAQAQPS 327
DB 29 GSSSQPHCPSPVSHQAQPHGQSLFADLSREELTAVMRFOTPLGGLVDAQAQPS 88
OY 328 DNCVPSVELQPLPPKAAALAHLDGSPPARALAIYFGRQOPNVSELYVGPPLPSPYM 387
DB 89 DNCIFSVELQPLPPKAAALAHLDGSPPARALAIYFGRQOPNVSELYVGPPLPSPYM 148
OY 388 RQVTERHRCGLPYHRRPVLFOEYLDIDQMFNRELPOASGLHCCFKKHGRNLVTMT 447
DB 149 RQVTERHRCGLPYHRRPVLFOEYLDIDQMFNRELPOASGLHCCFKKHGRNLVTMT 206
OY 448 TAPRGISGDRATWFGLYNYSIGAGFFLHHVGLLELNNKALDPAWMTIOKVYQGRYD 507
DB 207 APRGRISRERTWIGLYNYSIGAGFFLHHVGLLELNNKALDPAWMTIOKVYQGRYD 266
OY 508 SLAQLEAOFAGLVNVLIPDNGTGSMSLSPVPPGAPPLQFYQGRFVSQGRVAS 567
DB 267 DLGOLREKRSGLRLEVRVPLPPNGASSLSRNSPGLPPLQFSPGSOYVQGNLVYS 326
OY 568 SLWTFSEFGAFSGPRIFPVREGGERLYVEISLOEALATYGGNSPAMTTRYVDGFGMG 627
DB 327 SLWTFSEFGAFSGPRIFPVREGGERLYVEISLOEALATYGGNSPAMTTRYVDGFGMG 386
OY 628 KYTTPLRGVDCLYATYVDMHFLLESQAIRKIDAFCEVEONQGLPLRRHNSDLXSHYF 687
DB 387 RNSRGVLRGVDCYQATWMDIHILVKGAVQLLPGAVCFEENQGLPLRRHNSDLXSHYF 446
OY 688 GGIAGELVLYRKSSTILANTDYWDYFHPSGAIEIRFYATGYSSAFLEGATG- KYGNQ 745
DB 447 GGIAGELVLYRKSSTILANTDYWDYFHPSGAIEIRFYATGYSSAFLEGATG- KYGNQ 506
OY 746 VSEHTIGVHTSHAFKVDLDVAGLENNWAAEDMVFVPAWPSPEHOLQRLQVTRKLE 805
DB 507 VSEHTIGVHTSHAFKVDLDVAGLENNWAAEDMVFVPAWPSPEHOLQRLQVTRKLE 566
OY 806 MEQQAFLVGSATPRYLYLASNSHNKMGHPRGYRIQMLSEFAGEPLPONSSMARGFSWERY 865
DB 567 KEDLTAFSLGSLPRYLTYLASNOTNAMGHORG-----Y 599
OY 866 QLAVTOKEEPESSSVFNQNDPMAPTYVDSFINNETINGKDLVAAVTAAGFIHRAED 925
DB 600 QLAVTOKEEPESSSVFNQNDPMAPTYVDSFINNETINGKDLVAAVTAAGFIHRAED 659
OY 926 IPMTVTVNGVCGFFLPYNFEDDPSFYASDSTYFRGDODAGACEVPLCLPOAAACAP 985
DB 660 IPMTVTVNGVCGFFLPYNFEDDPSFYASDSTYFRGDODAGACEVPLCLPOAAACAP 719
OY 986 DLPFAFSHGCF 995
DB 720 DLPFAFSHGCF 729

RESULT 4
US-10-190-866A-1
; Sequence 1, Application US/10190866A
; Publication No. US20030092619A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim International GmbH

; TITLE OF INVENTION: Method for identifying substances which modulate
; TITLE OF INVENTION: Interleukin 4 signalling
; FILE REFERENCE: 1/1214 US
; CURRENT APPLICATION NUMBER: US/10/190,866A
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 60/306,552
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: EP 01117097.4
; PRIOR FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 401
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Description of Artificial Sequence: derivative of
; OTHER INFORMATION: STAT6 fragment
US-10-190-866A-1

Query Match 21.8%; Score 1165.5; DB 9; Length 401;
Best Local Similarity 68.7%; Pred. No. 3.4e-91;
Matches 248; Conservative 10; Mismatches 56; Indels 47; Gaps 10;

OY 38 SPLIGYWKIKGLVOPTRLLLEYEEKYEERDEGDKWRNKKFELGLEFPNLPYIDG 97
DB 2 SPLIGYWKIKGLVOPTRLLLEYEEKYEERDEGDKWRNKKFELGLEFPNLPYIDG 61
OY 98 DVKLTOSMALIRYIADKNNMGLGSPKERAISMLBGAVIDIRGVSRIRYSKDFELKYD 157
DB 62 DVKLTOSMALIRYIADKNNMGLGSPKERAISMLBGAVIDIRGVSRIRYSKDFELKYD 121
OY 158 FLTKLPEMKMEEDRLSHKTYLNGDVHTPDVFLYDALDVLYMDMCLDAPPKLYSFKK 217
DB 122 FLTKLPEMKMEEDRLSHKTYLNGDVHTPDVFLYDALDVLYMDMCLDAPPKLYSFKK 181
OY 218 RIFALPQIDKYLKSSKTYIAMPLOGWQATFGCGDHPKPSOSQELFQGG-----DGEPSQ 272
DB 182 RIFALPQIDKYLKSSKTYIAMPLOGWQATFGCGDHPKPSOSQELFQGG-----DGEPSQ 237
OY 273 LPHCPSPVSAQWTPHGSQSLFADLSREELTAVMRFOTPLGGLVDAQAQPSDNCVF 332
DB 238 LPHCPSPVSAQWTPHGSQSLFADLSREELTAVMRFOTPLGGLVDAQAQPSDNCVF 275
OY 333 SVELQPLPPKAAALAHLDGSPPARALAIYFGRQOPNVSELYVGPPLPSPYM 391
DB 276 SVELQPLPPKAAALAHLDGSPPARALAIYFGRQOPNVSELYVGPPLPSPYM 320
OY 392 V 392
DB 321 M 321

RESULT 5
US-09-844-988-4
; Sequence 4, Application US/09844988
; Patent No. US20020158764A1
; GENERAL INFORMATION:
; APPLICANT: Mercurio, Frank
; Zhu, Hengyi
; Barbosa, Miguel
; Li, Gfan
; Murray, Brian W.
; TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
; COMPLEX AND METHODS OF USE THEREFOR
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,988
FILING DATE: 26-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/910,820
FILING DATE: 1997-08-13

ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS: <unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-844-988-4

Query Match 21.7%; Score 1162; DB 9; Length 272;
Best Local Similarity 80.8%; Pred. No. 3.6e-91;
Matches 227; Conservative 4; Mismatches 32; Indels 18; Gaps 2;

QY 38 SPILGYWKIKGLVQPTLLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYIDG 97
DB 2 SPILGYWKIKGLVQPTLLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYIDG 61
QY 98 DVKLTSMAIIRYIADKHNMLGGSPKERAEISMLEGAVLDIRYGVSRAYSKDPELTAKVD 157
DB 62 DVKLTSMAIIRYIADKHNMLGGSPKERAEISMLEGAVLDIRYGVSRAYSKDPELTAKVD 121
QY 158 FLKSLPEMLKMFEDRLSHKTYLNGDHTHPDMLYDALDVLYMDPCLDAFPKLVCFK 217
DB 122 FLKSLPEMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPCLDAFPKLVCFK 181
QY 218 RIBAIPOIDKYLKSKXYIAMPLOGWQATFGGDHPKPSQSQALFOGGDGEPSQLPHCP 277
DB 182 RIBAIPOIDKYLKSKXYIAMPLOGWQATFGGDHPKPSDPRREFIVTDMAG-----VACL 235
QY 278 SVSPSAQPTWTHPGQSOLFADLSREELTAVNRFLTQRLGPEL 318
DB 236 GKTADADEMCDSGLSGLPDAAPG-----GPEL 264

RESULT 6

US-09-844-908-4
Sequence 4, Application US/09844908
Patent No. US20020151021A1
GENERAL INFORMATION:

APPLICANT: Mercurio, Frank

Zhu, Hengyi
Barbosa, Miguel
Li, Gian

TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,908
FILING DATE: 27-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997

ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS: <unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-844-908-4

Query Match 21.7%; Score 1162; DB 10; Length 272;
Best Local Similarity 80.8%; Pred. No. 3.6e-91;
Matches 227; Conservative 4; Mismatches 32; Indels 18; Gaps 2;

QY 38 SPILGYWKIKGLVQPTLLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYIDG 97
DB 2 SPILGYWKIKGLVQPTLLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYIDG 61
QY 98 DVKLTSMAIIRYIADKHNMLGGSPKERAEISMLEGAVLDIRYGVSRAYSKDPELTAKVD 157
DB 62 DVKLTSMAIIRYIADKHNMLGGSPKERAEISMLEGAVLDIRYGVSRAYSKDPELTAKVD 121
QY 158 FLKSLPEMLKMFEDRLSHKTYLNGDHTHPDMLYDALDVLYMDPCLDAFPKLVCFK 217
DB 122 FLKSLPEMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPCLDAFPKLVCFK 181
QY 218 RIBAIPOIDKYLKSKXYIAMPLOGWQATFGGDHPKPSQSQALFOGGDGEPSQLPHCP 277
DB 182 RIBAIPOIDKYLKSKXYIAMPLOGWQATFGGDHPKPSDPRREFIVTDMAG-----VACL 235
QY 278 SVSPSAQPTWTHPGQSOLFADLSREELTAVNRFLTQRLGPEL 318
DB 236 GKTADADEMCDSGLSGLPDAAPG-----GPEL 264

RESULT 7

US-09-844-988-6
Sequence 6, Application US/09844988
Patent No. US20020158764A1
GENERAL INFORMATION:

APPLICANT: Mercurio, Frank

Zhu, Hengyi
Barbosa, Miguel
Li, Gian

TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

Tue May. 20 13:19:00 2003

us-10-081-408-20.rapp

Page 5

```

1 COMPUTER READABLE FORM:
2
3 MEDIUM TYPE: floppy disk
4
5 COMPUTER: IBM PC compatible
6
7 OPERATING SYSTEM: PC-DOS/MS-DOS
8
9 SOFTWARE: PatentIn Release #1.0, Version #1.30
10
11 CURRENT APPLICATION DATA:
12
13 APPLICATION NUMBER: US/09/844,988
14
15 FILING DATE: 26-Apr-2001
16
17 CLASSIFICATION: <Unknown>
18
19 PRIOR APPLICATION DATA:
20
21 APPLICATION NUMBER: 08/910,820
22
23 FILING DATE: 1997-08-13
24
25 ATTORNEY/AGENT INFORMATION:
26
27 NAME: Makl, David J.
28
29 REGISTRATION NUMBER: 31,392
30
31 REFERENCE/DOCKET NUMBER: 860098.413C1
32
33 TELECOMMUNICATION INFORMATION:
34
35 TELEPHONE: (206) 622-4900
36
37 TELEFAX: (206) 682-6031
38
39 INFORMATION FOR SEQ. ID NO: 6:
40
41 SEQUENCE CHARACTERISTICS:
42
43 LENGTH: 272 amino acids
44
45 TYPE: amino acid
46
47 STRANDEDNESS: <Unknown>
48
49 TOPOLOGY: linear
50
51 SEQUENCE DESCRIPTION: SEQ ID NO: 6:
52
53 US-09-844-988-6

```

```

Query Match 21.7%: Score 1161: DB 9: Length 272:
Best Local Similarity 80.8%: Pred. No. 4.4e-91:
Matches 227: Conservative 4: Mismatches 32: Indels 18: Gaps 2:

QY 38 SPILGWKIKGLVQPTRLLEYLEEKYEENLYERDEGDKWRNKKELGLEPPLPYIDG 97
D 2 SPILGWKIKGLVQPTRLLEYLEEKYEENLYERDEGDKWRNKKELGLEPPLPYIDG 61
QY 98 DVKLQMSAIIIRYIADKNHNLGGSPKREAREISMLGAVLDIRYGSRIASQDFETLKYD 157
D 62 DVKLQMSAIIIRYIADKNHNLGGSPKREAREISMLGAVLDIRYGSRIASQDFETLKYD 121
QY 158 FLSLKPEMLKMFEDRLSHKTYLNGDHYTHPDEMLYDALDVLVYMDPMCLDAEPKLYSFKK 217
D 122 FLSLKPEMLKMFEDRLCHNTYLNQDHYTHPDEMLYDALDVLVYMDPMCLDAEPKLYSFKK 181
QY 218 RLEALPQIDKYIYKSSKXYIAMPQMGQATPGGDDHPKSSQSOEALFQGGGCGGESPQLPHCP 277
D 182 RLEALPQIDKYIYKSSKXYIAMPQMGQATPGGDDHPKSSPREFEITYTDMG-----VACL 235
QY 278 SVSPSAQPTWHPGSQLFADLSREELTAVMREFLQRLGPGIL 318
D 236 GKTADEDEWCADGALGSPDAAPG-----GGGL 264

RESULT 8
US-09-844-908-6
: Sequence 6, Application US/09844908
: Patent No. US20020151021A1
: GENERAL INFORMATION:
: APPLICANT: Mercutio, Frank
: 2hu, Hengyi
: Barbosa, Miguel
: Li, Gian
: Murray, Blion W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESS: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

```

```

? COMPUTER READABLE FORM:
? MEDIUM TYPE: floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
?
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/844,908
? FILING DATE: 27-Apr-2001
? CLASSIFICATION: <Unknown>
?
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/08/910,820
? FILING DATE: 12-AUG-1997
?
? ATTORNEY/AGENT INFORMATION:
? NAME: Makl, David J.
?
? REGISTRATION NUMBER: 31,392
? REFERENCE/DOCKET NUMBER: 860098.413C1
?
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (206) 622-4900
? TELEFAX: (206) 682-6031
?
? INFORMATION FOR SEQ. ID NO: 6:
?
? SEQUENCE CHARACTERISTICS:
? LENGTH: 272 amino acids
? TYPE: amino acid
? STRANDEDNESS: <Unknown>
? TOPOLOGY: linear
?
? SEQUENCE DESCRIPTION: SEQ ID NO: 6:
?
? US-09-844-908-6

```

```

Query Match      21.7% ; Score 1161; DB 10; Length 272;
Best Local Similarity 80.8% ; Pred. No. 4,4e-91;
Matches 227; Conservative 4; Mismatches 32; Indels 18; Gaps 2;

QY 38 SPILGYWKIKGLVOPTRLLEYLEEKEYEHELYERDEGKWKRNKKFELGLEPNNLPYIDG 97
Db 2 SPILGYWKIKGLVOPTRLLEYLEEKEYEHELYERDEGKWKRNKKFELGLEPNNLPYIDG 61
QY 98 DVKLTOSAAITRTYIADKNKNNMLGSGPKEAETISMLGAVLDIRYGVSRAYSXDFETLKYD 157
Db 62 DVKLTOSAAITRTYIADKNKNNMLGSGCKEAEISMLGAVLDIRYGVSRAYSXDFETLKYD 121
QY 158 FLSKLPEMLKMFEDRLSHKRTYLYNGDVTYHPDFMLYDALDVLYLXDPMCIDAFPKLVSEKK 217
Db 122 FLSKLPEMLKMFEDRLCHKRTYLYNGDVTYHPDFMLYDALDVLYLXDPMCIDAFPKLVSEKK 181
QY 218 RIEALPQIDKYLKSKSYIAMPLOGQATFFGGSDHPKSSOSQALPQGGDGEPSQLPHCP 277
Db 182 RIEALPQIDKYLKSKSYIAMPLOGQATFFGGSDHPKSSPREFIVTDNAG-----VACL 235
QY 278 SVSPSAQPTWTHPGQSOLFADLSREBLTAVMRFLTORLGPGL 318
Db 236 GKTADADEMCDAGLGPDAAPG-----GRL 264

RESULT 9
US-09-910-600-26
: Sequence 26, Application US/09910600
: Publication NO. US2003003631A1
: GENERAL INFORMATION:
: APPLICANT: Longphre, Mallinda
: APPLICANT: Chang, Han
: APPLICANT: Whitney, Gena
: TITLE OF INVENTION: NOVEL SIGLECS AND USES THEREOF
: FILE REFERENCE:: D0003NP
: CURRENT APPLICATION NUMBER: US/09/910,600
: PRIOR FILING DATE: 2001-07-20
: PRIOR APPLICATION NUMBER: 60/220,139
: PRIOR FILING DATE: 2000-07-21
: NUMBER OF SEQ ID NOS: 32
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 26
: LENGTH: 298
: TYPE: PRT
: ORGANISM: Artificial Sequence

```

```
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: L3cyto-Y641
; OTHER INFORMATION: alone
US-09-910-600-26

Query Match      21.7% Score 1160.5; DB 9; Length 298;
Best Local Similarity 89.3%; Pred. No. 5.6e-91;
Matches 221; Conservative 3; Mismatches 16; Indels 7; Gaps 1;

QY 38 SPILGYWKIKGLVQPTLLLEYLEEKKYEENHYERDEGDKMKNKFEELGLEPPNLPYYIDG 97
    |||||
DB 2 SPILGYWKIKGLVQPTLLLEYLEEKKYEENHYERDEGDKMKNKFEELGLEPPNLPYYIDG 61

QY 98 DVKLTQSMALIRYIADKHNMGSGPKERAETISMEGAVLDIRYGVSRVAYSKEDEFTLKVD 157
    |||||
DB 62 DVKLTQSMALIRYIADKHNMGSGPKERAETISMEGAVLDIRYGVSRVAYSKEDEFTLKVD 121

QY 158 FLSTLPEMLKAFEDRLSHKTYTLNGDHYTHPDPMYDALDVLYMDPMLDAFPKLVSEFK 217
    |||||
DB 122 FLSTLPEMLKAFEDRLSHKTYTLNGDHYTHPDPMYDALDVLYMDPMLDAFPKLVSEFK 181

QY 218 RIEAIPQIDKYLKSSKIYAMPLOGMOTFGGSDHPKSSQSEALFQGGDGEPSQLPHCP 277
    |||||
DB 182 RIEAIPQIDKYLKSSKIYAMPLOGMOTFGGSDHPKSSDLYPR-----GSPMSINVVP 234

QY 278 SVSPSAQ 284
    : ||
DB 235 TAGPLAQ 241

RESULT 10
US-10-081-408-5
; Sequence 5, Application US/10081408
; Patent No. US20020160482A1
; GENERAL INFORMATION:
; APPLICANT: Abrahams, n, Lars
; TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
; FILE REFERENCE: 13425-053001
; CURRENT APPLICATION NUMBER: US/10/081,408
; PRIOR APPLICATION NUMBER: SE 0100625-3
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 60/272,247
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Schistosoma japonicum
US-10-081-408-5

Query Match      21.6% Score 1159; DB 9; Length 218;
Best Local Similarity 100.0%; Pred. No. 4.6e-91;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 SPILGYWKIKGLVQPTLLLEYLEEKKYEENHYERDEGDKMKNKFEELGLEPPNLPYYIDG 97
    |||||
DB 2 SPILGYWKIKGLVQPTLLLEYLEEKKYEENHYERDEGDKMKNKFEELGLEPPNLPYYIDG 61

QY 98 DVKLTQSMALIRYIADKHNMGSGPKERAETISMEGAVLDIRYGVSRVAYSKEDEFTLKVD 157
    |||||
DB 62 DVKLTQSMALIRYIADKHNMGSGPKERAETISMEGAVLDIRYGVSRVAYSKEDEFTLKVD 121

QY 158 FLSTLPEMLKAFEDRLSHKTYTLNGDHYTHPDPMYDALDVLYMDPMLDAFPKLVSEFK 217
    |||||
DB 122 FLSTLPEMLKAFEDRLSHKTYTLNGDHYTHPDPMYDALDVLYMDPMLDAFPKLVSEFK 181

QY 218 RIEAIPQIDKYLKSSKIYAMPLOGMOTFGGSDHPK 254
    |||||
DB 182 RIEAIPQIDKYLKSSKIYAMPLOGMOTFGGSDHPK 218
```

```
RESULT 11
US-09-824-438-7
; Sequence 7, Application US/09824438
; Publication No. US20030073621A1
; GENERAL INFORMATION:
; APPLICANT: Butler, Jon P.
; APPLICANT: Hale, John E.
; APPLICANT: Heath Jr., William F.
; APPLICANT: Schoner, Brigitte E.
; APPLICANT: Helman, Mark L.
; APPLICANT: Becker, Gerald W.
; APPLICANT: Varshavsky, Alexander D.
; TITLE OF INVENTION: Beta-lipotropin and Uses Thereof
; FILE REFERENCE: X-12139
; CURRENT APPLICATION NUMBER: US/09/824,438
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GST/BLT fusion
US-09-824-438-7

Query Match      21.6% Score 1157.5; DB 9; Length 422;
Best Local Similarity 65.3%; Pred. No. 1.8e-90;
Matches 241; Conservative 17; Mismatches 68; Indels 43; Gaps 6;

QY 38 SPILGYWKIKGLVQPTLLLEYLEEKKYEENHYERDEGDKMKNKFEELGLEPPNLPYYIDG 97
    |||||
DB 2 SPILGYWKIKGLVQPTLLLEYLEEKKYEENHYERDEGDKMKNKFEELGLEPPNLPYYIDG 61

QY 98 DVKLTQSMALIRYIADKHNMGSGPKERAETISMEGAVLDIRYGVSRVAYSKEDEFTLKVD 157
    |||||
DB 62 DVKLTQSMALIRYIADKHNMGSGPKERAETISMEGAVLDIRYGVSRVAYSKEDEFTLKVD 121

QY 158 FLSTLPEMLKAFEDRLSHKTYTLNGDHYTHPDPMYDALDVLYMDPMLDAFPKLVSEFK 217
    |||||
DB 122 FLSTLPEMLKAFEDRLSHKTYTLNGDHYTHPDPMYDALDVLYMDPMLDAFPKLVSEFK 181

QY 218 RIEAIPQIDKYLKSSKIYAMPLOGMOTFGGSDHPKSSQSEALFQGGDGEPSQLPHCP 277
    |||||
DB 182 RIEAIPQIDKYLKSSKIYAMPLOGMOTFGGSDHPKS---DLVPRGSPGIHRDLVPR-- 236

QY 278 SVSPSAQPTWHPGQSOLFADLSREELTAVYKFLTORLPGIYDAQAQARPSDNCVFSVELQ 337
    |||||
DB 237 -----GSIEGRLENG-----QRLREG--DGPDPADGAGAGQADLE 270

QY 338 LPPKAAA-----LAHLDGSPPPAREALAVFFGRQOPNVS---ELVYGPLPHPS 365
    |||||
DB 271 HSLLYAAEKDEGPRYRMEHFRWGSPPDKRYGFMTEKSOPLVTLFKNAIIKINAYKKG 330

QY 386 YMRDVTVER 394
    ||::||
DB 331 EMRELIGOR 339

RESULT 12
US-09-824-438-6
; Sequence 6, Application US/09824438
; Publication No. US20030073621A1
; GENERAL INFORMATION:
; APPLICANT: Butler, Jon P.
; APPLICANT: Hale, John E.
; APPLICANT: Heath Jr., William F.
; APPLICANT: Schoner, Brigitte E.
; APPLICANT: Helman, Mark L.
; APPLICANT: Becker, Gerald W.
; APPLICANT: Varshavsky, Alexander D.
```


TITLE OF INVENTION: Beta-1ipotroin and Uses Thereof

FILE REFERENCE: X-12138
 CURRENT APPLICATION NUMBER: US/09/824,438
 CURRENT FILING DATE: 2001-10-29
 NUMBER OF SEQ ID NOS: 35
 SOFTWARE: Patent In Ver. 2.0
 SEQ ID NO 6
 LENGTH: 331
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: GST fusion
 OTHER INFORMATION: partner
 US-09-824-438-6

Query Match 21.6%; Score 1157; DB 9; Length 331;
 Best Local Similarity 72.1%; Pred. No. 1.3e-90;
 Matches 235; Conservative 8; Mismatches 43; Indels 40; Gaps 5;

QY 38 SPLIGWKIKGLVQPTRLLEYLEEYEEHLYERDEGDKWRNKKFELGLEFPNLPYIYG 97
 DB 2 SPLIGWKIKGLVQPTRLLEYLEEYEEHLYERDEGDKWRNKKFELGLEFPNLPYIYG 61
 QY 98 DVKLTOSMAIIRYIADKHNMLGGSPKERAISMLEGAVDIRGVSRISAKDFETLKYD 157
 DB 62 DVKLTOSMAIIRYIADKHNMLGGSPKERAISMLEGAVDIRGVSRISAKDFETLKYD 121
 QY 158 FLTKLPEMLKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVLVYMDPCIDAFPKLYSFKK 217
 DB 122 FLTKLPEMLKMFEDRLCHKTYLNGDHYTHPDFMLYDALDVLVYMDPCIDAFPKLYCFFKK 181
 QY 218 RIEAIPOIDKYLKSKYTIAPLOGMOATFGGSDHPKSSQSEALFQGGDGGESQLPKPC 217
 DB 182 RIEAIPOIDKYLKSKYTIAPLOGMOATFGGSDHPKSSQSEALFQGGDGGESQLPKPC 217
 QY 278 SVSPSAQPTWHPGSQLFADLSREELTAVMRFLTQRLGPDVDAQAARPDNCFVSVEIQ 337
 DB 237 -----GSIEGRELTC-----QRLRES--DGPGPADDGAGACADLE 270
 QY 338 LPPKAAA-----LAHLDRGSP 354
 DB 271 HSLVAAEKKEDEGPRMEHFRMGSP 296

RESULT 13

US-09-950-634-4
 Sequence 4, Application US/09950634
 Publication No. US20030032775A1
 GENERAL INFORMATION:

APPLICANT: Molnar-Kimber, Katherine L.
 Fallili, Amedeo F.
 Cagliano, Thomas J.
 Nakanishi, Koji
 Chen, Yanqiu

TITLE OF INVENTION: EFFECTOR PROTEINS OF RAPAMYCIN

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
 Dunner, L.L.P.
 STREET: 1300 I Street, N.W.
 CITY: Washington
 STATE: DC
 COUNTRY: USA
 ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/950,634
 FILING DATE: 13-Sep-2001
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/471,112
 FILING DATE: 06-JUN-1995
 APPLICATION NUMBER: US 08/384,524
 FILING DATE: 13-FEB-1995
 APPLICATION NUMBER: US 08/312,023
 FILING DATE: 26-SEP-1995
 APPLICATION NUMBER: US 08/207,975
 FILING DATE: 08-MAR-1994

ATTORNEY/AGENT INFORMATION:

NAME: Slekmann, Michael T.
 REGISTRATION NUMBER: 36,276
 REFERENCE/DOCKET NUMBER: 01142.0058-00000

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-408-4000
 TELEFAX: 202-408-4400

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1140 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear

MOLECULE TYPE: peptide
 SEQUENCE DESCRIPTION: SEQ ID NO: 4:
 US-09-950-634-4

Query Match 21.6%; Score 1155.5; DB 9; Length 1140;
 Best Local Similarity 92.8%; Pred. No. 1.3e-89;
 Matches 219; Conservative 1; Mismatches 11; Indels 5; Gaps 1;

QY 38 SPLIGWKIKGLVQPTRLLEYLEEYEEHLYERDEGDKWRNKKFELGLEFPNLPYIYG 97
 DB 2 SPLIGWKIKGLVQPTRLLEYLEEYEEHLYERDEGDKWRNKKFELGLEFPNLPYIYG 61
 QY 98 DVKLTOSMAIIRYIADKHNMLGGSPKERAISMLEGAVDIRGVSRISAKDFETLKYD 157
 DB 62 DVKLTOSMAIIRYIADKHNMLGGSPKERAISMLEGAVDIRGVSRISAKDFETLKYD 121
 QY 158 FLTKLPEMLKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVLVYMDPCIDAFPKLYSFKK 217
 DB 122 FLTKLPEMLKMFEDRLCHKTYLNGDHYTHPDFMLYDALDVLVYMDPCIDAFPKLYCFFKK 181
 QY 218 RIEAIPOIDKYLKSKYTIAPLOGMOATFGGSDHPKSSQSEALFQGGDGGESQLPKPC 217
 DB 182 RIEAIPOIDKYLKSKYTIAPLOGMOATFGGSDHPKSSQSEALFQGGDGGESQLPKPC 217

RESULT 14

US-09-844-988-3
 Sequence 3, Application US/09844988
 Patent No. US20020158764A1
 GENERAL INFORMATION:

APPLICANT: Mercurio, Frank
 Zhu, Hengyi
 Barbosa, Miguel
 Li, Glen
 Murray, Brian W.

TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED AND BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/950,634
 FILING DATE: 13-Sep-2001
 CLASSIFICATION: <Unknown>

```

1      APPLICATION NUMBER: US/09/844,988
2      FILING DATE: 26-Apr-2001
3      CLASSIFICATION: <Unknown>
4
5      PRIOR APPLICATION DATA:
6      APPLICATION NUMBER: 08/910,820
7      FILING DATE: 1997-08-13
8
9      ATTORNEY/AGENT INFORMATION:
10     NAME: Makl, David J.
11     REGISTRATION NUMBER: 31,392
12     REFERENCE/DOCKET NUMBER: 860098.413C1
13
14     TELECOMMUNICATION INFORMATION:
15     TELEPHONE: (206) 622-4900
16     TELEFAX: (206) 682-6031
17
18     INFORMATION FOR SEQ ID NO: 3:
19     SEQUENCE CHARACTERISTICS:
20     LENGTH: 283 amino acids
21     TYPE: amino acid
22     STRANDEDNESS: <Unknown>
23     TOPOLOGY: linear
24
25     SEQUENCE DESCRIPTION: SEQ ID NO: 3:
26     US-09-844-988-3

```

	Query Match	21.5%	Score 1154	DB 9	Length 282
	Best Local Similarity	93.6%	Pred. No. 1,9e-90		
	Matches	218	Conservative	3	Mismatches 8; Indels 4; Gaps 1;
QY	38	SPILGWWIKGLVOPFTLLLEYLEEKYEENLYERDEGDKWRNKKFELGLEPNNLPYYIDG	97		
DB	2	SPILGWWIKGLVOPFTLLLEYLEEKYEENLYERDEGDKWRNKKFELGLEPNNLPYYIDG	61		
QY	98	DVKLTQSAIIIRYIADKNINMLGGSPKEAEISMLGAVLDIRYGVSRITAYSKDFETLKYD	157		
DB	62	DVKLTQSAIIIRYIADKNINMLGGSPKEAEISMLGAVLDIRYGVSRITAYSKDFETLKYD	121		
QY	158	FLSKLPEMLKAFEDRLSHKTYTLNGDHYVHPFMTLDAYLVLYNPMCLDAPKLYSEFKK	217		
DB	122	FLSKLPEMLKAFEDRLSHKTYTLNGDHYVHPFMTLDAYLVLYNPMCLDAPKLYSEFKK	181		
QY	218	RLEAPQIDDKTLKSSKYYTAMPQCGQAATFGGGDHPKSSQSE---ALFQSGD	266		
DB	182	RLEAPQIDDKTLKSSKYYTAMPQCGQAATFGGGDHPKSSQSPRFITVDMQAAE	234		

RESULT 15
 US-09-844-988-5
 Sequence 5, Application US/09844988
 Patent No. US2002015876A1
 GENERAL INFORMATION:
 APPLICANT: Mercurio, Frank
 Zhu, Hengyi
 Barbosa, Miguel
 Li, Gian
 Murray, Brian W.
 TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
 COMPLEX AND METHODS OF USE THEREFOR
 NUMBER OF SEQUENCES: 25
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED AND BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 City: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/844,988
 FILING DATE: 26-Apr-2001
 CLASSIFICATION: <unknown>
 PRIOR APPLICATION DATA:

```

: APPLICATION NUMBER: 08/910,820
: FILING DATE: 1997-08-13
: ATTORNEY/AGENT INFORMATION:
: NAME: Mak1, David J.
: REGISTRATION NUMBER: 31,392
: REFERENCE//DOCKET NUMBER: 860098.413C1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 283 amino acids
: TYPE: amino acid
: STRANDEDNESS: <unknown>
: TOPOLOGY: linear
: SEQUENCE DESCRIPTION: SEQ ID NO: 5:
: SS-09-844-988-5

```

	Query Match	21.5%;	Score 1154;	DB 9;	Length 282;
	Best Local Similarity	93.6%;	Pred. No. 1.9e-90;		
	Matches 218;	Conservative 3;	Mismatches 8;	Indels 4;	Gaps 1.
QY	38	SPILGYWKIKGLVQPTRLLEYLEEKEYEENLYERDEGDKWRNKKFELGLEPNNLPYYIDG	97		
Db	2	SPILGYWKIKGLVQPTRLLEYLEEKEYEENLYERDEGDKWRNKKFELGLEPNNLPYYIDG	61		
QY	98	DVKLTQSAIIIRYIADKRNHMLGGSPKEEAELSMLEGAVIDIYGVSRIAYSDFELKYD	157		
Db	62	DVKLTQSAIIIRYIADKRNHMLGGSPKEEAELSMLEGAVIDIYGVSRIAYSDFELKYD	121.		
QY	158	FLSKLPEMLKMFEDRLSHKTYLNGDHVHPDPEMLDALDVLVLYNPMPCIDAPPKLVSEFK	217		
Db	122	FLSKLPEMLKMFEDRLSHKTYLNGDHVHPDPEMLDALDVLVLYNPMPCIDAPPKLVSEFK	181		
QY	218	RIEAIPOIDIKYLKSSSKYIAMPLOGMOATFGGSDHPKPSOSOE----	ALFQGD 266		
Db	182	RIEAIPOIDIKYLKSSSKYIAMPLOGMOATFGGSDHPKPSOSOE----	ALFQGD 234		

Search completed: May 20, 2003, 12:02:26
Job time : 163.01 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:09:54 ; Search time 38.5066 seconds
(without alignments)
2491.578 Million cell updates/sec

Title: US-10-081-408-20

Perfect score: 5357

Sequence: 1 MDWLRNLLFLMAAQAQINMA.....QAAACAPDLPAFSGHGFESH 998

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3967	74.1	763	2 JCS234	amine oxidase (cop
2	3326	62.1	762	2 A54411	amine oxidase (cop
3	1398.5	26.1	746	2 S34656	amine oxidase (cop
4	1394	26.0	751	2 A54053	amine oxidase (cop
5	1140	21.3	219	2 A26484	glutathione transf
6	985	18.4	218	2 A45556	glutathione S-tran
7	922	17.2	218	2 A45523	glutathione transf
8	663	12.4	209	2 A48388	glutathione S-tran
9	489.5	9.1	218	2 S33860	glutathione transf
10	489.5	9.1	218	2 A29794	glutathione transf
11	486.5	9.1	218	2 S32425	glutathione transf
12	485.5	9.1	218	2 S65674	glutathione transf
13	484.5	9.0	218	2 A39375	glutathione transf
14	483.5	9.0	218	2 A47486	glutathione transf
15	481.5	9.0	217	2 JX0095	glutathione transf
16	476.5	8.9	218	2 S13202	glutathione transf
17	476.5	8.9	218	2 A23732	glutathione transf
18	475.5	8.9	218	2 B34159	glutathione transf
19	473.5	8.8	218	2 S01719	glutathione transf
20	471.5	8.8	218	2 S01719	glutathione transf
21	469.5	8.8	218	2 A29036	glutathione transf
22	468.5	8.7	218	2 B29231	glutathione transf
23	468.5	8.7	218	2 XURG4	glutathione transf
24	460.5	8.6	218	2 A46143	mu-class glutathio
25	460.5	8.6	218	2 B28946	glutathione transf
26	452.5	8.4	225	2 A35295	glutathione transf
27	437.5	8.2	220	2 S18464	glutathione transf
28	413	7.7	735	2 B41856	amine oxidase (fla
29	397.5	7.4	757	2 E64889	amine oxidase (cop

30	395.5	7.4	219	2 S50146	major allergen Dpi
31	377	7.0	684	2 A56102	amine oxidase (cop
32	368.5	6.9	660	2 G90330	amine oxidase (cop
33	363	6.8	648	2 A48646	amine oxidase (cop
34	357.5	6.7	756	2 T48139	copper amine oxida
35	353.5	6.6	650	2 G71412	probable amine oxi
36	340	6.3	638	2 JCS234	phenylethylamine o
37	325.5	6.1	674	2 UC7251	amine oxidase (cop
38	324	6.0	687	2 T47403	amine oxidase-like
39	320	6.0	674	2 C44239	amine oxidase (cop
40	305	5.7	759	2 E84854	probable copper am
41	300	5.6	660	2 AH2234	copper amine oxida
42	295	5.5	692	2 S04963	amine oxidase (cop
43	260.5	4.9	142	2 S17462	glutathione transf
44	253	4.7	210	2 S71958	glutathione transf
45	251.5	4.7	208	2 S41933	glutathione transf

ALIGNMENTS

RESULT 1

JCS234

amine oxidase (copper-containing) (EC 1.4.3.6) AOC2 precursor - human

C:Species: Homo sapiens (man)

C:Date: 13-Mar-1997 #sequence-revision 13-Mar-1997 #text-change 18-Feb-2000

C:Accession: JCS234

R:Zhang, X.; McIntyre, W.S.

Gene 179, 279-286, 1996

A:Title: Cloning and sequencing of a copper-containing, topa quinone-containing mon

A:Reference number: JCS234; MUID:97128319; PMID:8972912

A:Accession: JCS234

A:Contents: Placenta

A:Molecule type: mRNA

A:Residues: 1-763 <ZHA>

A:Cross-references: GB:U39447; NID:91399031; PIDN:AAC50919.1; PID:91399032

C:Comment: This enzyme catalyzes the oxidation of primary amines to the correspondi

Ca(II) and 1 molecule of covalently-bound topa quinone. It also can scavenge clrc

C:Genetics:

A:Gene: GDB:AOC2; DAO2

A:Cross-references: GDB:4562632

A:Map position: 17q21-17q21

C:Superfamily: amiloride-binding protein

C:Keywords: copper; metalloprotein; oxidoreductase; quinoprotein; topaquinone

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-763/Product: amine oxidase (copper-containing) #status predicted <MAT>

F:444-520/Binding site: copper (His) #status predicted

F:471/Modified site: topaquinone (Tyr) #status predicted

Query Match 74.1% Score 3967; DB 2; Length 763;

Best local similarity 99.9%; Pred. No. 3.5e-293;

Matches 735; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	263	QGGGGGSPQLPHCPSPSPNAQPTWHPGQSLFADLSREELTAYMRELTORLGLVDAA	322
DB	28	RGGGGGSPQLPHCPSPSPNAQPTWHPGQSLFADLSREELTAYMRELTORLGLVDAA	87
QY	323	QARSDNCVSVELQLPPKAAALAHLDGSPPARALAIYFGRQOPNVSELVGPPLP	382
DB	88	QARSDNCVSVELQLPPKAAALAHLDGSPPARALAIYFGRQOPNVSELVGPPLP	147
QY	383	HPSYMRDVTYERHGGPLPYHRRPVLFQYLDIDIMIFNRELPOASGLLHCCFYKHGRN	442
DB	148	HPSYMRDVTYERHGGPLPYHRRPVLFQYLDIDIMIFNRELPOASGLLHCCFYKHGRN	207
QY	443	LVTMTTPRGLQSDRATWGLYINISAGFLHHVGLLNNKALDPAWTIOXYFQ	502
DB	208	LVTMTTPRGLQSDRATWGLYINISAGFLHHVGLLNNKALDPAWTIOXYFQ	267
QY	503	GRYDLSAQLAEPAQGLVNVVLLIPDNGTGSMSLSKSPVPGAPAPLQFYDQGRFVSOG	562
DB	268	GRYDLSAQLAEPAQGLVNVVLLIPDNGTGSMSLSKSPVPGAPAPLQFYDQGRFVSOG	327

Qy	563	SRVASLWTFEEFGAASGPPIFQVRFQGEELVYEISLOEALITYGNSPAAWTTFRVDS	622
Dd	328	SRVASLWTFEEFGAASGPPIFQVRFQGEELVYEISLOEALITYGNSPAAWTTFRVDS	387
Qy	623	GFQMGKYYTPTLRGVDPPYATLVYDMHFLESOAPKTIIDAPCFVEQNOGLPLRRHSDI	682
Dd	388	GFQMGKYYTPTLRGVDPPYATLVYDMHFLESOAPKTIIDAPCFVEQNOGLPLRRHSDI	447
Qy	663	YSHFEGGLAEVLVLRKSKSTLNTDYMOTVYFHPSCAIEIRFATGITYSSAFIFGATGKY	742
Dd	448	YSHFEGGLAEVLVLRKSKSTLNTDYMOTVYFHPSCAIEIRFATGITYSSAFIFGATGKY	507
Qy	743	GNQVSEHTLGVTHSHAFKVDLDVAGLEMMVNAEDVAFVPMVPMSPBEHQDRLQVTRK	802
Dd	508	GNQVSEHTLGVTHSHAFKVDLDVAGLEMMVNAEDVAFVPMVPMSPBEHQDRLQVTRK	567
Qy	803	LLEMEQAAFLVGSATPRYLYLASNSHKNKGHPGXYRIOMLSFAGEFLPONSSMARGFWS	862
Dd	568	LLEMEQAAFLVGSATPRYLYLASNSHKNKGHPGXYRIOMLSFAGEFLPONSSMARGFWS	627
Qy	863	ERYOLAVQRKEEPPSSSVFNQNDPAPATYVDFSDFINNETIACKDLVAAWTTAGFLIIPH	922
Dd	628	ERYOLAVQRKEEPPSSSVFNQNDPAPATYVDFSDFINNETIACKDLVAAWTTAGFLIIPH	687
Qy	923	AEDIPNTVYVNGVGFLLRPYNFEDEDEPSFYSAISITYFRGDQDAGACVNPPLACLPDAAA	982
Dd	688	AEDIPNTVYVNGVGFLLRPYNFEDEDEPSFYSAISITYFRGDQDAGACVNPPLACLPDAAA	747
Qy	983	CAPDLPATSHGCGFSHN 998	
Dd	748	CAPDLPATSHGCGFSHN 763	

RESULT 2

amine oxidase (protein-containing) (EC 1.4.3.6), serum, precursor - bovine
C.Species: Bos primigenius taurus (cattle)
C.Date: 09-Sep-1994 #sequence_revision 06-Feb-1995 #text_change 18-Feb-2000
C.Accession: A54411; B38081; A48242; S65408
R.MU; D.; Medzhitnadszky, K.F.; Adams, G.W.; Mayer, P.; Hines, W.M.; Burlingame, A.L.; Sm
J. Biol. Chem. 269, 9926-9932, 1994
A.Title: Primary structures for a mammalian cellular and serum copper amine oxidase.
A.Reference number: A54411; MUID:94193686; PMID:8144587
A.Accession: A54411
A.Molecule type: mRNA }
A.Residues: 1-762 <MU>
A.Cross-references: GB:S65983; NID:9546215; PIDN:AB30397.1; PID:9546216
R.MU; D.; James, S.M.; Smith, A.J.; Brown, D.E.; Dooley, D.M.; Kilman, J.P.
J. Biol. Chem. 267, 7979-7982, 1992
A.Title: Tyrosine codon corresponds to topa quinone at the active site of copper amine c
A.Reference number: A38081; MUID:92235001; PMID:1569055
A.Accession: B38081
A.Molecule type: Protein
A.Residues: 463-465 'D', 467-473, 'X', 475-485 <MU2>
R.James, S.M.; MU, D.; Wemmer, D.; Smith, A.J.; Kaur, S.; Maltby, D.; Burlingame, A.L.;
Science 248, 981-987, 1990
A.Title: A new redox cofactor in eukaryotic enzymes: 6-hydroxydopa at the active site of
A.Reference number: A48242; MUID:90260648; PMID:2111581
A.Accession: A48242
A.Status: preliminary
A.Molecule type: protein
A.Residues: 468-472 <JAN>
Eur. J. Biochem. 237, 93-99, 1996
A.Title: Half-of-the-sites reactivity of bovine serum amine oxidase. Reactivity and chem
A.Reference number: S65408; MUID:96203913; PMID:8620899
A.Accession: S65408
A.Molecule type: protein
A.Residues: 463-469, 'X', 471-487 <DEB>
C.Superfamily: anion-binding protein
C.Keywords: copper; glycoprotein; oxidoreductase; quinoprotein; topaquinone
E.I-16/Domain: signal sequence #status predicted <SI>

F.117-763/Product: amine oxidase (copper-containing) #status predicted <MAT>
F.136,21,665/Binding site: carboxylate (Asn) (covalent) #status predicted
F.470/Modified site: topaquinone (Tyr) #status experimental

Query Match	62.1%; Score 3326; DB 2; Length 762;
-------------	--------------------------------------

Matches 606; Conservative 57; Mismatches 70; Indels 8; Gaps 1;

QY 263 QGGDGE-----PSQLPHCPSVSPSAQPWTHPGOSQLFADLSREELTAVMRELTORL 314

Db 19 EGGVGSEEGVGKQCHPSLPSPSPSPDQPTWTHPDQSLFADLSREELTYVMSFLTQQL 78

QY 315 GPGLVDAQARPSDNCVFSEVELQLPPKAAALAHLDRGSPPPAREALAI VFFGRQPQNV 374

Db 79 GPDLVDAQRPSCNCFSEVLEQLPPKAAALAHIDRGSPPPAREALAIVFEGGPQPNVT 138

375 ELVGP LPHPSYMRDVTVERHGGLPYHRRPVLFQEYLDIDQMFNREL PQASGLHHCC 434

Db 139 ELVGP L P O P S Y M R D V T V E R H G G P L P Y R R P V L L R E Y L D I D O M I F N R E L P Q A G V L H H C C 198

QY 435 FYKHRGNLVMTTAPRGLOS GDRATWFGLYYNISGAGFFELHHVGLLELVN HKALDPARW 494

Db 199 SYKGGOKLLTMSAPRGVOSGDRSTWFGIYYNITKGGPYLHPVGELELLVDHKALDPADW 258

0V 495 TIOKVEYGRYDLSLAOLEAOFEAGLVNVLIPDNGTGSWSLKSPPVPGAPPIQFYPO 554

Db 359 TVOKVEEGRYENTAOIEFOEAGOVNMYTPDGTGFWSLKSOYPPGPTPIQEHPO 318

07 555 CBPESVOCSPVASI WTESEFCI CAESCPIEDVPBOGEPRI VYETSIQFAI ATYGNSPAA 614

[illegible]

674 5 680 685 690 695 700 705 710 715 720 725 730 735 740 745 750 755 760 765 770 775 780 785 790 795 800 805 810 815 820 825 830 835 840 845 850 855 860 865 870 875 880 885 890 895 900 905 910 915 920 925 930 935 940 945 950 955 960 965 970 975 980 985 990 995

[illegible]

THE UNIVERSITY OF CHICAGO

— — — — —

[illegible][illegible]

22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051

[illegible]

DD 339 UKLUV I KANDEL E E Q H A I F L G S A S F K I D I L A S N U S N A W O H F A O I A N I Q I V S E A O O F M F U N S O L

855 SMARGFWEKIQLA VIKRAEEPPSSSVFNQNDPMAFI VDFSDI INNEI IAGNDLVAMVI 51

DB 019 PMEKAFSWMGRIQLAITQKKEIEPSSSVFNQNDPMIPIVDESDFINNEI IAGNDLVAMVI 0/

915 AGFLHPHAEDIPNIVGVNGVGF LKPI NFD EDEPSFI SADSII FKBDQDAGACEVNP L 9/

Db 679 AGFLHPHAEDIPNTVTGNGVGFFLRPNFEDQEPSMDSADSIYFREGQDAGSCEINPL / 31

QY	975	ACLPQAAACAPDLP	AFSHGGF	995
----	-----	----------------	---------	-----

RESULT 3

amino oxidase (copper-containing) (EC 1.4.3.6), kidney, precursor - rat
 N:Alternative names: amiloride-binding protein, long form
 N:Contains: amiloride-binding protein, short form
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 13-Jan-1995 #sequence revision 13-Jan-1995 #text_change 18-Feb-2000
 R:Accession: S36847; S36848; I51904; S34656; S34657
 R:Rindupella, E.; Renard, S.; Volley, N.; Waldmann, R.; Chassande, O.; Lazdunski, M.
 Eur. J. Biochem. 216, 679-687, 1993
 A:Title: Molecular cloning and functional expression of different molecular forms of
 A:Reference number: S36847; MUID:93587321; PMID:8375402

A:Accession: S36847
A:Molecule type: mRNA
A:Residues: 1-746 <LIN2>
A:Cross-references: EMBL:X73911; NID:g395064; PIDN:CAA52116.1; PID:g395065
A:Accession: S36848
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 529-746 <LIN2>
A:Cross-references: EMBL:X73912; NID:g395066; PIDN:CAA52117.1; PID:g395067
R:Verity, K.; Fuller, P.J.
Am. J. Physiol. 266, C1505-C1512, 1994
A:Title: Isolation of a rat amiloride-binding protein cDNA clone: tissue distribution and
A:Reference number: 151904; MUID:94295632; PMID:8023885
A:Accession: 151904
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 530-586, 'T', 588-632 <VER>
A:Cross-references: GB:S70383; NID:g547214; PIDN:AAB31157.1; PID:g547215
A:Note: In Genbank entry S70383, release 116.0, the source is designated as Rattus sp.
C:Superfamily: amiloride-binding protein
C:Keywords: alternative initiators; oxidoreductase
F:1-746/Product: amine oxidase (copper-containing), kidney, precursor #status predicted
F:529-746/Product: amiloride-binding protein short form #status predicted <SHORT>

Query Match 26.1%; Score 1398.5; DB 2; Length 746;
Best Local Similarity 40.3%; Pred. No. 7.1e-98;
Matches 294; Conservative 127; Mismatches 266; Indels 43; Gaps 13;

OY 291 OSOLFADLSREELTAVMRFLTORLGPGLVDAOARPSDCVSEVELQPLPKAALAHLDNR 350
DB 29 KARFFADLSQELKAVHSFLMNRBELGLOPSKEPTLAKNSVLEMLPKKHVLLKLE 88
OY 351 GSPPPAREALAIYFFGQPPQPNVSELVYGLPHPSYMRDYVERHGGPLRYHRRPVLEQ 410
DB 89 GRKGPNEARAVIFFGQDPNPVTEFAVGLPRPYTRALS-PPRGHLLSWSSPSIAE 147
OY 411 YLIDDMIFNRELQASGLAH-----CCFYHNRGNLYTMTARGLSGDRATWFG 463
DB 148 Y----DLLYH-TLKRATMPRLHQFLDYGSEFLCCDDRCFLTDVAPRGVASSGGRSWEI 202
OY 464 LYVNSAGPFLHNVGELVNHKALDPAWTKQVFOGQYUUSLQKQFAGLVNV 523
DB 203 VQRYVE--GIFLHPTGELILDHGSTVDQMRVQLMTNGAFYNNPEELAKYAVGEVD 260
OY 524 VLIDP---NGTGSMSLSPY-----PPGAPRLQYPOGFRFSVQSGSRVASS 568
DB 261 VLLEDPLPNGTE-----KPLFSSYKPRGEFHTPVNVAGPVGSGRYLLEGWTVLG 315
OY 569 LMTSFGIGAFSPRIEDVRFQGERLYEISLOALAIYGGNSPAAATRTTYVGGFGMK 628
DB 316 GMSFSYLRSSSGIQIFENVLGGERVAYEVSQEAVALYGGHTPAGMKOTKIDYDVGWGLS 375
OY 629 YTTLTGVDGPPYATVYDMHFLLESQAPKTIIPAFCVEQNOGLRLRHSDYS--H 685
DB 376 VTHELAPGIDCPETATFIDAFHYIDSQGVHYPHMLCFEMPTGVLRLKHNSNFKGFPN 435
OY 686 YFGCLAEIVLVVRSMTLLNTDYVMDYFHPGSAIEIFYATGYISSAFLGATGKYNQ 745
DB 436 FYAGIKGVVILVLTSTYVNDYIMDFIFYSNGVEMAKMHAATGYHAHFYREGLRHGR 495
OY 746 VSEHTLGVTHSAHFQVDLVAGLENNVMAEDVYFVMAVPMSFHOLOLRQVTRKLL 805
DB 496 LQTHLGNIRHVLHYRVMDVAGTKNSFQTLJMKLEMLJTNWSPSHSLYQPTLQOTYXS 555
OY 806 MEOAALVGSATPRYLTLASNSKMGHPRGYRIOMLSFAGEPLPONSMSYARGFSEWERY 865
DB 556 OEHQAAEFPGOTLPKYLFLFSSPOKCKWGRKSTYRQIHSMAQVLPQGOEERAVTWAY 615
OY 866 QLAATQREKEEPPSSSVENQNDPNAPTVDFSDFI--NNETIAGKDLVAVWTAGFLIHPAE 924
DB 616 PLAVTKYRESEKYSSTLVNQNQNDPMDPPVYFEELKNNENIEDDLVAVWTAGFLIHPSE 675
OY 925 DIPNTVTVNGVGVFLRPYNFEDDEBPSYSADSIYFRGDQDAGACEVNPVLACLPPAACA 984

DB 676 DVNNTATPGNSGVFLRPNFPEDEPSLASRDIVYV-WPQDKGLNVQ--RWLPEDRRCL 732
OY 985 PDLPAFSHGG 994
DB 733 VS-PFPSYNG 741

RESULT 4
A54053
amine oxidase (copper-containing) (EC 1.4.3.6) ABP1 precursor - human
N:Alternate names: amiloride-binding protein; diamine oxidase; histaminase
C:Species: Homo sapiens (man)
C:Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 18-Feb-2000
C:Accession: A54053; A54410; I38708; A38276; S42495
R:Chassande, O.; Renard, S.; Barbry, P.; Lazdunski, M.
J. Biol. Chem. 269, 14484-14489, 1994
A:Title: The human gene for diamine oxidase, an amiloride binding protein. Molecula
A:Reference number: A54053; MUID:94237856; PMID:8182053
A:Accession: A54053
A:Molecule type: DNA
A:Residues: 1-751 <CHA>
A:Cross-references: EMBL:X78212; NID:g463242; PIDN:CAA5046.1; PID:g463243
R:Novotny, W.F.; Chassande, O.; Baker, M.; Lazdunski, M.; Barbry, P.
J. Biol. Chem. 269, 9921-9925, 1994
A:Title: Diamine oxidase is the amiloride-binding protein and is inhibited by amilo
A:Reference number: A54410; MUID:94193685; PMID:8144586
A:Accession: A54410
A:Molecule type: Protein
A:Residues: 20-27, 'A', 29-39 <NOV>
A:Cross-references: PIDN:AAB30395.1; PID:g546213
A:Experimental source: placenta
A:Note: Sequence extracted from NCBI backbone (NCBI:P:146046)
R:Zhang, X.; Kim, J.; McIntire, W.S.
Biochem. Genet. 33, 261-268, 1995
A:Title: cDNA sequences of variant forms of human placenta diamine oxidase.
A:Reference number: I38708; MUID:96113540; PMID:8595053
A:Accession: I38708
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-331, 'F', 333-751 <RES>
A:Cross-references: EMBL:U11862; NID:g535353; PIDN:AAC50270.1; PID:g535356
R:Barbry, P.; Champé, S.; Chassande, O.; Munemitsu, S.; Champigny, G.; Lingueglia,
Proc. Natl. Acad. Sci. U.S.A. 87, 7347-7351, 1990
A:Title: Human kidney amiloride-binding protein: cDNA structure and functional exp
A:Reference number: A38276; MUID:91017502; PMID:2217167
A:Accession: A38276
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-265, 'K', 267-751 <K>
A:Cross-references: GB:M55602; GB:M36335; NID:g87653; PIDN:AAA58358.1; PID:g17796
A:Note: This sequence has been corrected in reference A54053
C:Genetics:
A:Gene: GDB:ABP1
A:Cross-references: GDB:127105; OMIM:104610
A:Map position: 7q31-7q32
A:Introns: 524/1; 619/2; 663/3
C:Function:
A:Description: oxidation of putrescine and histamine
C:Superfamily: amiloride-binding protein
C:Keywords: homodimer; oxidoreductase; quinoprotein; topaquinone
F:1-19/Dominant: signal sequence #status predicted <SIG>
F:461/Modified site: topaquinone (Tyr) #status predicted

Query Match 26.0%; Score 1394; DB 2; Length 751;
Best Local Similarity 39.8%; Pred. No. 1.6e-97;
Matches 293; Conservative 124; Mismatches 281; Indels 38; Gaps 11;

OY 285 PWTBGSOLFADLSREELTAVMRFLTORLGPGLVDAOARPSDCVSEVELQPLPKAALAHLDNR 344
DB 23 PGTLPRAKGVFSDLSQELKAVHSFLMNRBELGLOPSKEPTLAKNSVLEMLPKKHVLLKLE 82
OY 345 LAHNRGSPPARREALAIYFFGQPPQPNVSELVYGLPHPSYMRDYVERHGGPLRYHRR 404

Query Match 21.3%; Score 1140; DB 2; Length 219;
Best Local Similarity 98.6%; Pred. No. 4.4e-79;
Matches 213; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

39 PILGVWKIKGLVQPRRLLEVEENKYLEERDGDGKWRKKFELGLEFPNLPYYIOGD 98

Db	4	PILGQYWKJNGVQPRLLLEYLEEYEBHLYERDEGGDKMRKKFELGLEPPLNLPYIDGO	63
QY	99	VKLQMSMAIIIRYIADKHNMLGGSPKEREAEISMLGAVLIDIRGYSRATYSKDEPTELVDF	158
Db	64	VKLQMSMAIIIRIADKHNMLGGCPKEREAEISMLGAVLIDIRGYSRATYSKDEPTELVDF	133
QY	159	LSKLEPMLKMEFDRLSHNTYTLNGDHYVHPDMLDYALDVLVLYMDPMLDAFAFKLYSEFKR	218
Db	124	LSKLEPMLKMEFDRLCRHTYTLNGDHYVHPDMLDYALDVLVLYMDPMLDAFAFKLYSEFKR	183
QY	219	IEAIPQIDKLYLSSKXYIAMPVLOGQAFGGGDHPK	254
Db	184	IEAIPQIDKLYLSSKXYIAMPVLOGQAFGGGDHPK	219

RESULT 6
A45556
glutathione S-transferase - fluke (Schistosoma mansoni)
C:Species: Schistosoma mansoni
C:Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 03-May-1996
C:Accession: A45556
R:Wright, M.D.; Harrison, R.A.; Welder, A.M.; Newport, G.R.; Mitchell, G.F.
Mol. Biochem. Parasitol. 49, 177-179, 1991
A:Title: Another 26-kilodalton glutathione S-transferase of Schistosoma mansoni
A:Reference number: A45556; MUID:92131046; PMID:1775156
A:Accession: A45556
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-218 <MR>
A:Note: sequence extracted from NCBI backbone (NCBIF:77814)
C:Superfamily: glutathione transferase

	Query Match	18.4%	Score 985	DB 2	length 218
	Best Local Similarity	82.9%	Pred. No. 2,7e-67		
	Matches 180	Conservative 19	Mismatches 18	Indels 0	Gaps
QY	38	SPILGYSKIKGLVQPTRLILEYEEKYEBHLIYERDEBDCKMRNKKFELGTFEPMLPYIIDG	97		
Db	2	APKLYGMYKIKGLVQPTRLILEYEEKYEBHLIYERDEBDCKMRNKKFELGTFEPMLPYIIDG	61		
QY	98	DVKLTQSAIIIRYIADCKNNMLGSGPKREAEISMEGAVLDIRGVSIAYSKDPEITLKYD	157		
Db	62	DVKLTQSAIIIRYIADCKNNMLGSGPKREAEISMEGAILDIRGVSIAYSKDPEITLKYD	121		
QY	158	FLSKLPEMLKMFEDRLSHKTYLNGDHTHPDPEMLYDALDVLYMDPQCILDAPFKLYSFKK	217		
Db	122	FLNCPFGMLKMFEEBRLSHNTYLNLDGDKTHTPPEMLYDALDVLYMDPQCILDAPFKLYSFKK	181		
QY	218	RTEALPQIDKYLKSSKYAMPLOQWQMTFFGGGDHPK	254		
Db	182	RTEALPPIKNTLNSNRYTKMPLQGSMTFFGGGDHPK	218		

RESULT 7
A45523
glutathione transferase (EC 2.5.1.18) - fluke (Schistosoma mansoni)
C:Species: Schistosoma mansoni
C:Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 18-Jun-1999
C:Accession: A45523; A44941
R:Rottstein, F.; Kieny, M.P.; Verwaerde, C.; Torpier, G.; Pierce, R.J.; Balloul, J.M.;
Mol. Biochem. Parasitol. 41, 35-44, 1990
A:Title: Molecular cloning and tissue distribution of a 26-kilodalton Schistosoma man
A:Reference number: A45523; MUID:90348716; PMID:2385266
A:Accession: A45523
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-218 <TR>
A:Cross-references: GB:M1106; NID:g161004; PIR:AAA2888.1; PID:g552241
R:Henkle, K.J.; Davern, K.M.; Wright, M.D.; Ramos, A.O.; Mitchell, G.F.
Mol. Biochem. Parasitol. 40, 23-34, 1990
A:Title: Comparison of the cloned genes of the 26- and 28-kilodalton glutathione S-tr
A:Reference number: A44941; MUID:90271935; PMID:1693415

Db 64 IDGSHKKTQSNALIRYLARKHHLDOETFEERIRADIVENQVMDFPMQIMLCYNDFEKO 123
 QY 155 KVDLSKLPKMLKMFEDRLSHKTYLNGDHYTHPPDMLYDALDVLYMDPKCLDAFPKLYS 214
 Db 124 KPEFLKTIPEKMKLYSEFLGRPFAGDKYTVYDFLAYDILDQYMEFPKCLDAFPNLRD 183
 QY 215 FKRIEAIPOIDKYLKSSKYIAMP 239
 Db 184 FLARFEGIKKISAYMKSSRYLSTPI 208

RESULT 10

A29794
 N:Alternate names: glutathione S-transferase Yb1; 11gadin
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 28-Aug-1989 #sequence-revision 28-Aug-1989 #text-change 18-Jun-1999
 C:Accession: A29794; A25510; A24085; B61004; F24735; A33397; S27111; S17167; A26187
 R:Chang, C.; Saitzman, A.G.; Sorensen, N.S.; Hlupkaka, R.A.; Liso, S.
 J. Biol. Chem. 262, 11901-11903, 1987
 A:Title: Identification of glutathione S-transferase Yb-1 mRNA as the androgen-repressed
 A:Reference number: A29794; MUID:87308179; PMID:3040722
 A:Accession: A29794
 A:Molecule type: mRNA
 A:Residues: 1-218 <CHA>
 A:Cross-references: GB:J02810; NID:9204514; PIDN:AAA41293.1; PID:9204515
 R:Lai, H.C.J.; Grove, G.; Tu, C.P.D
 Nucleic Acids Res. 14, 6101-6114, 1986
 A:Title: Cloning and sequence analysis of a cDNA for a rat liver glutathione S-transferase
 A:Reference number: A25510; MUID:86312882; PMID:2875437
 A:Accession: A25510
 A:Molecule type: mRNA
 A:Residues: 1-218 <LAT>
 A:Cross-references: GB:X04229; NID:956337; PIDN:CAA27811.1; PID:956338
 R:Ding, G.U.F.; Lu, A.Y.H.; Pickett, C.B.
 J. Biol. Chem. 260, 13268-13271, 1985
 A:Reference number: A24085; MUID:86033768; PMID:3840477
 A:Accession: A24085
 A:Molecule type: mRNA
 A:Residues: 1-198, 'NC', 201-218 <DIN>
 A:Cross-references: GB:M11719; NID:9204502; PIDN:AAA41287.1; PID:9204503
 A:Experimental source: clone pGTA/C44
 R:Chang, L.H.; Hsieh, J.C.; Chen, W.L.; Tam, M.F.
 Electrophoresis 11, 589-593, 1990
 A:Title: Identification of rat liver glutathione S-transferase Yb subunits by partial N-
 local isoelectric focusing gel
 A:Reference number: A61004; MUID:91031411; PMID:2226415
 A:Accession: B61004
 A:Molecule type: protein
 A:Residues: 2-26 <CH2>
 R:Mennervik, B.; Alln, P.; Guttenberg, C.; Jansson, H.; Tahr, M.K.; Warholm, M.; Jornvad
 Proc. Natl. Acad. Sci. U.S.A. 82, 7202-7206, 1985
 A:Title: Identification of three classes of cytosolic glutathione transferase common to
 A:Reference number: A24735; MUID:86042634; PMID:3864155
 A:Accession: F24735
 A:Molecule type: protein
 A:Residues: 2-20 <MAN>
 R:Hsieh, J.C.; Liu, L.F.; Chen, W.L.; Tam, M.F.
 Biochem. Biophys. Res. Commun. 162, 1147-1154, 1989
 A:Title: Expression of Yb-1 glutathione S-transferase using a baculovirus expression sys
 A:Reference number: A33397; MUID:89350924; PMID:2669745
 A:Accession: A33397
 A:Molecule type: protein
 A:Residues: 2-21, 212-218 <HS1>
 R:Katusz, R.M.; Bono, B.; Colman, R.F.
 Arch. Biochem. Biophys. 298, 667-677, 1992
 A:Title: Identification of Tyr(115) labeled by S-(4-bromo-2,3-dioxobutyl)glutathione in
 A:Reference number: S27111; MUID:93037509; PMID:1416995
 A:Accession: S27111
 A:Molecule type: protein
 A:Residues: 2-20, 83-86, 'X', 88-96, 109-115, 'X', 117-122 <KAT>
 R:Hsieh, J.C.; Huang, S.C.; Chen, W.L.; Lai, Y.C.; Tam, M.F.

Biochem. J. 278, 293-297, 1991
 A:Title: Cysteine-86 is not needed for the enzymic activity of glutathione S-transferase
 A:Reference number: S17167; MUID:91354218; PMID:1863338
 A:Accession: S17167
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 2-218 <HS2>
 C:Superfamily: glutathione transferase
 C:Keywords: dimer; transferase

Query Match 9.18; Score 486.5; DB 2; Length 218;
 Best Local Similarity 44.48; Pred. No. 2.2e-29;
 Matches 94; Conservative 38; Mismatches 70; Indels 5; Gaps 1;

QY 40 ILGWYKIKGLVOPRLLEYLEEKYEHLERDEG----DKWNRKFELECPNLPY 94
 Db 4 ILGWYNRGLTHPIRLLELYTSSYEKKRYAMGADPYDRSQMLNEKFKGLDPPNLPYL 63
 QY 95 IDGDKLTQSNALIRYLADKRNMLGSGPKERAESMDEGAVLDIRYGVSRAYSKDEFTL 154
 Db 64 IDGSRKKTQSNALIRYLARKHHLGTEFEERIRADIVENQVMDFPMQIMLCYNDFEKO 123
 QY 155 KVDLSKLPKMLKMFEDRLSHKTYLNGDHYTHPPDMLYDALDVLYMDPKCLDAFPKLYS 214
 Db 124 KPEFLKTIPEKMKLYSEFLGRPFAGDKYTVYDFLAYDILDQYHIFEPKCLDAFPNLRD 183
 QY 215 FKRIEAIPOIDKYLKSSKYIAMP 239
 Db 184 FLARFEGIKKISAYMKSSRYLSTPI 208

RESULT 11

S32425
 N:Alternate names: glutathione transferase mu4
 C:Species: Homo sapiens (man)
 C:Date: 07-Apr-1994 #sequence-revision 07-Apr-1994 #text-change 18-Jun-1999
 C:Accession: S32425; S29337
 R:Zhong, S.; Spurr, N.K.; Hayes, J.D.; Wolf, C.R.
 Biochem. J. 291, 41-50, 1993
 A:Title: Deduced amino acid sequence, gene structure and chromosomal location of a no
 A:Reference number: S32424; MUID:93228631; PMID:8471052
 A:Accession: S32425
 A:Molecule type: DNA
 A:Residues: 1-218 <ZHO>
 A:Cross-references: EMBL:X68677; NID:931934; PIDN:CAA48637.1; PID:931935
 A:Note: The authors translated the codon AAG for residue 182 as Arg and CCA for resid
 C:Genetics:
 A:Gene: GDB:GSTM4
 A:Cross-references: GDB:134191; OMIM:138333
 A:Map position: 1p13.3-1p13.3
 A:Insertions: 12/3; 38/1; 59/3; 87/1; 120/3; 152/3; 189/3
 C:Superfamily: glutathione transferase
 C:Keywords: dimer; transferase

Query Match 9.18; Score 486.5; DB 2; Length 218;
 Best Local Similarity 44.18; Pred. No. 2.2e-29;
 Matches 94; Conservative 38; Mismatches 76; Indels 5; Gaps 1;

QY 40 ILGWYKIKGLVOPRLLEYLEEKYEHLERD-----EGDKNRKFELECPNLPY 94
 Db 4 ILGWYDRGLAHARLLLETTDSYEKKRTTGGAPDYDSQMLNEKFKGLDPPNLPYL 63
 QY 95 IDGDKLTQSNALIRYLADKRNMLGSGPKERAESMDEGAVLDIRYGVSRAYSKDEFTL 154
 Db 64 IDGAKKTQSNALIRYLARKHHLGTEFEERIRADIVENQVMDFPMQIMLCYNDFEKO 123
 QY 155 KVDLSKLPKMLKMFEDRLSHKTYLNGDHYTHPPDMLYDALDVLYMDPKCLDAFPKLYS 214
 Db 124 KPEFLKTIPEKMKLYSEFLGRPFAGDKYTVYDFLAYDILDQYHIFEPKCLDAFPNLRD 183
 QY 215 FKRIEAIPOIDKYLKSSKYIAMPLOGMATFG 247

A:Residues: 1-218 <CO2>
 A:Cross-references: GB:M66234; NID:9306818; PIDN:AAA57347.1; PID:9306819
 A:Experimental source: HeLa cells
 R:Ross, V.L.; Board, P.G.
 Biochem. J. 294, 373-380, 1993
 A:Title: Molecular cloning and heterologous expression of an alternatively spliced human
 A:Reference number: S56782; MUID:93384505; PMID:8373352
 A:Accession: JX0095
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-218 <ROS>
 A:Cross-references: EMBL:M99422
 A:Experimental source: testis
 R:Comstock, K.E.; Wierstgen, M.; Hao, X.Y.; Henner, W.D.; Mannervik, B.
 Arch. Biochem. Biophys. 311, 487-495, 1994
 A:Title: A comparison of the enzymatic and physicochemical properties of human glutathione
 A:Reference number: S45684; MUID:94263230; PMID:8203914
 A:Contents: annotation
 R:Taylor, J.B.; Oliver, J.; Sherrington, R.; Pemble, S.E.
 Biochem. J. 274, 587-593, 1991
 A:Title: Structure of human glutathione S-transferase class Mu genes.
 A:Reference number: I37438; MUID:91174774; PMID:2006920
 A:Accession: J37438
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 39-120 <RES>
 A:Cross-references: EMBL:X56837; NID:931936; PIDN:CAA40167.1; PID:931937
 A:Genetics:
 A:Gene: GDB:GSTM4
 A:Cross-references: GDB:134191; OMIM:138333
 A:Map position: 1p13.3-1p13.3
 A:Introns: 12/3; 38/1; 59/3; 87/1; 120/3; 152/3; 189/3
 C:Superfamily: glutathione transferase
 C:Keywords: alternative splicing; dimer; transferase

Query Match 9.0%; Score 483.5; DB 2; Length 218;
 Best Local Similarity 43.7%; Pred. No. 3.7e-29;
 Matches 94; Conservative 38; Mismatches 78; Indels 5; Gaps 1;

QY 38 SPILGYMKIKGLVOPTRLLLEYLEKEYEHLERDEG-----DKWRNKKFELGLEFPNLP 92
 DB 2 SMTGGMWDIRGLAHLRLLEYTDSYEKKYTMGDADPDYRSQWLNKFKLGLDFPNLP 61

QY 93 YIIDGVKLTQSMALIRIADKHNHNLGSGPKERAISMLEGAVLDIRGVSRIVASKDFE 152
 DB 62 YLIDGAHKITQSMALICTIARKHNLGSETEEKIRVDLENQAMDVSNOCLARVCSPDFE 121

QY 153 TLKVDFTLSKLPDMLKMFEDRLSHKTYLNGDHTHPDMLYDALDVLYMDPWCIDAPFKL 212
 DB 122 KIKREYLEELPTMQHFQFLGKRPMFVGDKITFVDFLAYDVLDRIFERNCLDAFPNL 181

QY 213 VSFKKRIEAIPOIDKYLKSSKRYIAMPLOGMOATFG 247
 DB 182 KDFISREGLEKISAYMKSRLPLPYTRVAVMG 216

RESULT 15
 JX0095
 glutathione transferase (EC 2.5.1.18) b - guinea pig
 N:Alternate names: glutathione S-alkyltransferase; glutathione S-aryltransferase; glutathione
 C:Species: Cavia porcellus (guinea pig)
 C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 18-Jun-1993
 C:Accession: JX0095
 R:Kamel, K.; Oshino, R.; Hara, S.
 J. Biochem. 107, 111-117, 1990
 A:Title: Amino acid sequence of glutathione S-transferase b from guinea pig liver.
 A:Reference number: JX0095; MUID:90236961; PMID:2332413
 A:Accession: JX0095
 A:Molecule type: protein
 A:Residues: 1-217 <KAM>
 C:Comment: Glutathione transferases are a family of enzymes that are multifunctional in
 C:Superfamily: glutathione transferase
 C:Keywords: transferase

Query Match 9.0%; Score 481.5; DB 2; Length 217;
 Best Local Similarity 45.1%; Pred. No. 5.2e-29;
 Matches 92; Conservative 35; Mismatches 72; Indels 5; Gaps 1;

QY 41 LGYWMKIKGLVOPTRLLLEYLEKEYEHLERDEG-----DKWRNKKFELGLEFPNLPYI 95
 DB 4 LGYWMIRGLTHPIRLILEYTNNGSEKKRYNMGDADPDYRSQWLNKFKLGLDFPNLPYI 63

QY 96 DGDVLTQSMALIRIADKHNHNLGSGPKERAISMLEGAVLDIRGVSRIVASKDFETIK 155
 DB 64 DGTNKLQSMALIRIADKHNHNLGSGPKERAISMLEGAVLDIRGVSRIVASKDFETIK 123

QY 156 VDFLSKLPDMLKMFEDRLSHKTYLNGDHTHPDMLYDALDVLYMDPWCIDAPFKLVSF 215
 DB 124 AEFLEGIRPKMLFQFQFLGKRPMFVGDKITFVDFLAYDVLDRIFERNCLDAFPNL 183

QY 216 KRIEAIPOIDKYLKSSKRYIAMP 239
 DB 184 ISFEGLEKISSYMKSSRLPKPL 207

Search completed: May 20, 2003, 11:39:29
 Job time : 42.5066 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 08:02:28 ; Search time 23.557 Seconds
(without alignments)
1757.161 Million cell updates/sec

Title: US-10-081-408-20

Perfect score: 5357
Sequence: 1 MDWLRLFLMAAASINVA.....QAACAPDLPFSGGFSIN 998

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3967	74.1	763	1 AOC3_HUMAN	Q16883 homo sapien
2	3365	62.8	765	1 AOC3_MOUSE	O70423 mus musculu
3	3338	62.1	762	1 AOCX_BOVIN	O46406 bos taurus
4	3326	62.1	762	1 AOCX_BOVIN	O29437 bos taurus
5	2556	47.7	756	1 AOC2_HUMAN	O75106 homo sapien
6	1398.5	26.1	746	1 ABP_RAT	P36633 ratlus norv
7	1394	26.0	751	1 ABP_HUMAN	P19801 homo sapien
8	1315	24.5	321	1 AOC3_RAT	O08590 ratlus norv
9	1144	21.4	218	1 GT26_SCHJA	P08515 schistosoma
10	985	18.4	218	1 GT27_SCHMA	P35661 schistosoma
11	922	17.2	218	1 GT26_SCHMA	P15964 schistosoma
12	719.5	13.4	220	1 GT29_FASHE	P56598 fasciola he
13	708	13.2	217	1 GT27_FASHE	P31670 fasciola he
14	690	12.9	217	1 GT28_FASHE	P31671 fasciola he
15	689	12.9	217	1 GT26_FASHE	P30112 fasciola he
16	489.5	9.1	217	1 GTM1_MOUSE	P10649 mus musculu
17	489.5	9.1	217	1 GTM1_RAT	P04905 ratlus norv
18	485.5	9.1	217	1 GTM2_RABIT	P46409 oryctolagus
19	485.5	9.1	218	1 GTM6_MOUSE	O35660 mus musculu
20	484.5	9.0	217	1 GTM2_HUMAN	P28161 homo sapien
21	483.5	9.0	218	1 GTM4_HUMAN	O03013 homo sapien
22	481.5	9.0	217	1 GTM2_HUMAN	P16413 cavia porce
23	476.5	8.9	217	1 GTM2_CAVPO	O00285 cricetus
24	476.5	8.9	217	1 GTM2_MESAU	P30116 mesocricetu
25	475.5	8.9	217	1 GTM2_MOUSE	P15636 mus musculu
26	473.5	8.8	217	1 GTM1_HUMAN	P09488 homo sapien
27	471.5	8.8	217	1 GTM5_MOUSE	P46439 homo sapien
28	469.5	8.8	217	1 GTM2_HUMAN	P08009 ratlus norv
29	463.5	8.7	217	1 GTM2_RAT	P08010 ratlus norv
30	460.5	8.6	217	1 GTM3_MOUSE	P12126 mus musculu
31	455.5	8.5	224	1 GTM3_HUMAN	P48774 mus musculu
32	443.5	8.3	224	1 GTM5_MOUSE	P20136 gallus gall
33	440.5	8.2	219	1 GTM2_CHICK	

34	413	7.7	755	1 AMO_KIEAE	P49250 klebsiella
35	397.5	7.4	757	1 AMO_ECOLI	P46883 escherichia
36	395.5	7.4	219	1 GTM1_DERPT	P46419 dermatophag
37	377	7.0	683	1 AMO1_ARTGO	O59118 artroobacte
38	363	6.8	648	1 AMO2_ARTS1	O07121 artroobacte
39	363	6.8	648	1 AMO2_ARTS1	O07123 artroobacte
40	340	6.3	638	1 PAOX_ARTGO	P46881 artroobacte
41	320	6.0	674	1 AMO_PEA	O43077 pisum sativ
42	311.5	5.8	666	1 AMO_LENCU	P49252 lens culina
43	310	5.8	671	1 AMO1_ASPNG	O12556 aspergillus
44	295.	5.5	692	1 AMO_PICAN	P12807 picchia angu
45	256.5	4.8	208	1 GTP_DIRIM	P46426 dirofilaria

ALIGNMENTS

RESULT 1	
AOC3_HUMAN	STANDARD: PRT: 763 AA.
AC	O16853:
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	Membrane copper amine oxidase (EC 1.4.3.6) (Vascular adhesion protein-1) (VAP-1) (HPAO).
GN	AOC3 OR VAP1.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX	NCBI_Taxid=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Placenta;
RX	MEDLINE=97128319; PubMed=8972912;
RA	Zhang X., McIntire W.S.;
RT	"Cloning and sequencing of a copper-containing, topaquinone-
RT	containing monamine oxidase from human placenta.";
RL	Gene 179:279-286(1996).
RN	[2]
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC	TISSUE=Lung;
RX	MEDLINE=98317014; PubMed=9653080;
RA	Smith D.J., Salmi M., Bono P., Hellman J., Lau T., Jalkanen S.;
RT	"Cloning of vascular adhesion protein 1 reveals a novel
RT	multifunctional adhesion molecule.";
RL	J. Exp. Med. 188:17-27(1998).
CC	-1- FUNCTION: CELL ADHESION PROTEIN THAT PARTICIPATE IN LYMPHOCYTE
CC	RECIRCULATION BY MEDIATING THE BINDING OF LYMPHOCYTES TO
CC	PERIPHERAL LYMPH NODE VASCULAR ENDOTHELIAL CELLS IN AN L-SELECTIN-
CC	INDEPENDENT FASHION. HAS A MONAMINE OXIDASE ACTIVITY.
CC	-1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +
CC	H(2)O(2).
CC	-1- COFACTOR: Binds 1 copper ion and 1 topaquinone per subunit.
CC	-1- SUBUNIT: HOMODIMER.
CC	-1- SUBCELLULAR LOCATION: Type II membrane protein.
CC	-1- TISSUE SPECIFICITY: MOST STRONGLY EXPRESSED ON THE HIGH
CC	ENDOTHELIAL VENULES OF PERIPHERAL LYMPH NODES AND ON HEPATIC
CC	ENDOTHELIA.
CC	-1- PTM: Topaquinone (TPQ) is generated by copper-dependent
CC	autooxidation of a specific tyrosyl residue (by similarity).
CC	-1- PTM: N- AND O-GLYCOSYLATED.
CC	-1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC	*****
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	*****
DR	EMBL: U039447; AAC50919.1; -.

```

DR EMBL: AF067406; AAC25170.1;
DR Genem: HGNC:550; AOC3.
DR MIM: 603735;
DR InterPro: IPR000269; CuNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
DR Pfam: PF02727; Cu_amine_oxid2; 1.
DR Pfam: PF02728; Cu_amine_oxid3; 1.
DR PRINTS: PR00766; CUDAOXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
DR Oxioreductase; Copper; TPO; Glycoprotein; Transmembrane;
KW Signal-anchor; Cell adhesion; Polymorphism; Metal-binding.
FT DOMAIN 1 5
FT TRANSMEM 6 26
FT (POTENTIAL).
FT (EXTRACELLULAR (POTENTIAL)).
FT MOD_RES 471 471 TOPAOUINONE (BY SIMILARITY).
FT METAL 520 520 COPPER (POTENTIAL).
FT METAL 522 522 COPPER (POTENTIAL).
FT METAL 684 684 COPPER (POTENTIAL).
FT BINDING 673 673 AMLOXIDE (BY SIMILARITY).
FT CARBOHYD 137 137 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 592 592 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 618 618 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 666 666 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 317 317 Y -> H (IN DBSNP:438287).
FT SEQUENCE 763 AA; 84621 MW; 58AD55605EC9D228 CRC64;

Query Match 74.1%; Score 3967; DB 1; Length 763;
Best Local Similarity 99.9%; Pred. No. 3.5e-289;
Matches 735; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 263 QGGDGGSPSLPHCPSPVSPAQPTWHPGQSOLFADLSREELTAVMRFLTORGLGVDA 322
DB 28 RCGGSGSPSLPHCPSPVSPAQPTWHPGQSOLFADLSREELTAVMRFLTORGLGVDA 87
OY 323 QARSDNCVSVELQLEPKKAALAHDRGSPPRARELATIYEFRRQOPNVSELVGP 382
DB 88 QARSDNCVSVELQLEPKKAALAHDRGSPPRARELATIYEFRRQOPNVSELVGP 147
OY 383 HPSYMRDVYERHGGPLPYHRRPPLFOEYLDIDIMFNRELPOASGLLHHCFCFKHGRN 442
DB 148 HPSYMRDVYERHGGPLPYHRRPPLFOEYLDIDIMFNRELPOASGLLHHCFCFKHGRN 207
OY 443 LVMTTAPRGLQSGDRATWFGLYYINISGAGFLLHVGLELLVNHKALDPAKMTIQKVFYQ 502
DB 208 LVMTTAPRGLQSGDRATWFGLYYINISGAGFLLHVGLELLVNHKALDPAKMTIQKVFYQ 267
OY 503 GRYYDSLALQLEAGFELVNVVLIIPDNGTGGSMLSKPVPPGPARPLQFYQGRFVSVOG 562
DB 268 GRYYDSLALQLEAGFELVNVVLIIPDNGTGGSMLSKPVPPGPARPLQFYQGRFVSVOG 327
OY 563 SRVASSLMTSPFGIAGFSGRIEFDVRFQGERLYVEISLQELATLYGNSPAAMTTRYVDG 622
DB 328 SRVASSLMTSPFGIAGFSGRIEFDVRFQGERLYVEISLQELATLYGNSPAAMTTRYVDG 387
OY 623 GFQMGKTTPLTRGVOCPTYATVYDMHFLLESQAPKTRIDACFVEFQNOGLPLRRHNSDL 682
DB 388 GFQMGKTTPLTRGVOCPTYATVYDMHFLLESQAPKTRIDACFVEFQNOGLPLRRHNSDL 447
OY 683 YSHYFGGLAEFVLVYVRSMTLLNVDYMDVYHPSGAIETRFVYGTYSNLFEGAGKXY 742
DB 448 YSHYFGGLAEFVLVYVRSMTLLNVDYMDVYHPSGAIETRFVYGTYSNLFEGAGKXY 507
OY 743 GNOVSEHTLGTVTHSHAFKVDLDVAGLEKVMVAEDMVFYPMAYPSPREHOLQRLQYTRK 802
DB 508 GNOVSEHTLGTVTHSHAFKVDLDVAGLEKVMVAEDMVFYPMAYPSPREHOLQRLQYTRK 567
OY 803 LEMEBEAAALVGSATPRYLYLASNHSNKGHPGRIYIOMLSFGEPLPONSSNARCFSSW 862

```

```

DB 568 LEMEBEAAALVGSATPRYLYLASNHSNKGHPGRIYIOMLSFGEPLPONSSNARCFSSW 627
OY 863 ERYOLAVTORKEEPPSSSVENQNDPMAPTYDVSDFINNETIACKDLVAMWTAGFLHHPH 922
DB 628 ERYOLAVTORKEEPPSSSVENQNDPMAPTYDVSDFINNETIACKDLVAMWTAGFLHHPH 687
OY 923 AEDIPNTVTVNGVGFILRPYNNFDEDEPSYSDSIYFRGQDAGACEVNPDLAPQAAA 982
DB 688 AEDIPNTVTVNGVGFILRPYNNFDEDEPSYSDSIYFRGQDAGACEVNPDLAPQAAA 747
OY 983 CAPDLPAFSGHGFSSHN 998
DB 748 CAPDLPAFSGHGFSSHN 763

RESULT 2
AOC3_MOUSE
ID AOC3_MOUSE STANDARD; PRT; 765 AA.
AC 070423;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Membrane copper amine oxidase (EC 1.4.3.6) (Vascular adhesion protein-1) (VAP-1).
GN AOC3 OR VAP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid-10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/c, and 129/SvJ;
RX MEDLINE-98414290; PubMed-9743358;
RA Bono P., Salmi M., Smith D.J., Leppanen I., Horelli-Kultunen N.,
RA Palotie A., Jalkanen S.;
RT Isolation, structural characterization, and chromosomal mapping of
RT the mouse vascular adhesion protein-1 gene and promoter."
RL J. Immunol. 161:2953-2960(1998).
CC -1- FUNCTION: CELL ADHESION PROTEIN THAT PARTICIPATE IN LYMPHOCYTE
CC RECIRCULATION BY MEDIATING THE BINDING OF LYMPHOCYTES TO
CC PERIPHERAL LYMPH NODE VASCULAR ENDOTHELIAL CELLS IN AN L-SELECTIN-
CC INDEPENDENT FASHION. HAS A MONOMINE OXIDASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
CC H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
CC -1- PTM: Topaquinine (TPO) is generated by copper-dependent
CC autoxidation of a specific tyrosyl residue (By similarity).
CC -1- PTM: N- AND O-GLYCOSYLATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF054831; AAC23747.1;
DR EMBL: AF078705; AAC35839.1;
DR MIM: MGI:1306797; AOC3.
DR InterPro: IPR000269; CuNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
DR Pfam: PF02727; Cu_amine_oxid2; 1.
DR Pfam: PF02728; Cu_amine_oxid3; 1.
DR PRINTS: PR00766; CUDAOXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
DR Oxioreductase; Copper; TPO; Glycoprotein; Transmembrane;
KW Signal-anchor; Cell adhesion; Metal-binding.
FT DOMAIN 1 6
FT CYTOPLASMIC (POTENTIAL).

```

Query Match
Best Local Similarity 83.0%; Score 3365; DB 1; Length 765;
Matches 610; Conservative 62; Mismatches 63; Indels 0; Gaps 0;

FT	TRANSMEM	7	27	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT	DOMAIN	28	765	(POTENTIAL)
FT	MOD_RES	471	471	TOPAQUINONE (BY SIMILARITY).
FT	METAL	520	520	COPPER (POTENTIAL).
FT	METAL	522	522	COPPER (POTENTIAL).
FT	METAL	684	684	COPPER (POTENTIAL).
FT	BINDING	673	673	AMILORIDE (BY SIMILARITY).
FT	CARBOHYD	137	137	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	232	232	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	294	294	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	592	592	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	659	659	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	666	666	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	765 AA;	84533 MW;	74896D67D3DBB44D CRC64;

Query Match
Best Local Similarity 83.1%; Score 3338; DB 1; Length 762;
Matches 616; Conservative 42; Mismatches 75; Indels 8; Gaps 1;

FT <th>TRANSMEM</th> <th>7</th> <th>27</th> <th>SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)</th>	TRANSMEM	7	27	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT	DOMAIN	28	765	(POTENTIAL)
FT	MOD_RES	471	471	TOPAQUINONE (BY SIMILARITY).
FT	METAL	520	520	COPPER (POTENTIAL).
FT	METAL	522	522	COPPER (POTENTIAL).
FT	METAL	684	684	COPPER (POTENTIAL).
FT	BINDING	673	673	AMILORIDE (BY SIMILARITY).
FT	CARBOHYD	137	137	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	232	232	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	294	294	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	592	592	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	659	659	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	666	666	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	762 AA;	84883 MW;	BB43D04776744AF2 CRC64;

```

OY 315 GPGVDAQAARPSDNCVSEVLEQLPKKAAALAHLDKSGSPPARALAVFEGGQOPRVN 374
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 79 GPDVDAQAARPSDNCVSEVLEQLPKKAAALAHLDKSGSPPARALAVFEGGQOPRVN 138
OY 375 ELVVGPLPSPYMDVYERHGGPLPYRRHVPVLEQVLDIDQMFNRELPAAGLLHHC 434
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 139 ELVVGPLPSPYMDVYERHGGPLPYRRHVPVLEQVLDIDQMFNRELPAAGLLHHC 198
OY 435 FYKRGRLVMTTAPRGLSGDRATWGLYNTISGAFGLHHVGLLVNKKALDPARW 494
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 199 SYKGGGGLVMTTAPRGLSGDRATWGLYNTISGAFGLHHVGLLVNKKALDPARW 258
OY 495 TIQVYFQGRYDLSLAOLEAFEGVGLVNVVLIIPNGTSGMSLSPVPPGAPPLQFPQ 554
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 259 TIQVYFQGRYDLSLAOLEAFEGVGLVNVVLIIPNGTSGMSLSPVPPGAPPLQFPQ 318
OY 555 GPRSVGSRAVSSLMWTFSGLAFSGPRIDVRFQGERLYEISLOALAIYGSNPPA 614
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 319 GTRSVGSRAVSSLMWTFSGLAFSGPRIDVRFQGERLYEISLOALAIYGSNPPA 378
OY 615 MTTVYVGGFCMGKYTPPLTRGVDCPYLATYVDHMFLLSQAPKTIRACVFEQNGLP 674
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 379 MLTRIMDCFEWGFAPPLTRGVDCPYLATYVDHMFLLSQAPKTIRACVFEQNGLP 438
OY 675 LRRHSDLSHYFGLAETVVLVRSMSSTLNVYMDTVFHPSCAIEIRFATGYSSAF 734
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 439 LRRHSDLSHYFGLAETVVLVRSMSSTLNVYMDTVFHPSCAIEIRFATGYSSAF 498
OY 735 LFGATGKYGNVSEHTLCTGTHSAHFVLDVAGLENNVAAEDVAFVMAVPMSPHQ 794
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 499 FFGTAOKXGNGVRENTLCTGTHSAHFVLDVAGLENNVAAEDVAFVPTVPMSPHQ 558
OY 795 ORLOVTRKLEMEBOAFLVSATPRYLXLSNHNKMGHRCGRYIOMLSTAGEPLPNS 854
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 559 ORLOVTRKLEMEBOAFLVSATPRYLXLSNHNKMGHRCGRYIOMLSTAGEPLPNS 618
OY 855 SMARFSEWERYOLAVTORKEEEPSSSVFNONDPAPTVDSDFINNTETIGKDLVAVT 914
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 619 STERAIMSGRQLAVTORKETEPSSSVFNONDPAPTVDSDFINNTETIGKDLVAVT 678
OY 915 AGFLHPIPAEDIPNTVYVNGVGFLLRPYNFDEDPSTYSADSIYFRGDQAGACVNP 974
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 679 AGFLHPIPAEDIPNTVYVNGVGFLLRPYNFDEDPSTYSADSIYFRGDQAGACVNP 738
OY 975 ACLPOAAACAPDLPAFSHGGF 995
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 739 ACLPOAAACAPDLPAFSHGGF 759

```

RESULT 4
AOCX_BOVIN
ID AOCX_BOVIN STANDARD: PRT: 762 AA.

```

AC 029437;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Copper amine oxidase, liver isozyme precursor (EC 1.4.3.6) (Amine
DE oxidase [copper-containing]) (Serum amine oxidase) (SAO).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Liver.
RX MEDLINE=94193686; Pubmed=8144587;
RA Mu D., Medzhradszky K.F., Adams G.W., Mayer P., Hines W.M.,
RA Burlingame A.L., Smith A.J., Cal D., Kilman J.P.;
RT "Primary structures for a mammalian cellular and serum copper amine
RT oxidase";
RL J. Biol. Chem. 269:9926-9932(1994).

```

```

RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=93090748; Pubmed=1457410;
RA James S.M., Palcic M.M., Scaman C.H., Smith A.J., Brown D.E.,
RA Dooley D.M., Mure M., Kilman J.P.;
RT "Identification of topaquinine and its consensus sequence in copper
RT amine oxidases";
RL Biochemistry 31:12147-12154(1992).
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +
CC H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
CC -1- ENZYME REGULATION: INHIBITED BY AMILORIDE IN A COMPETITIVE MANNER.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: Liver.
CC -1- PTM: Topaquinine (TPQ) is generated by copper-dependent
CC autooxidation of a specific tyrosyl residue (by similarity).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, S69583; AAB30397.1; -;
DR EMBL, L27218; AAA30525.1; -;
DR InterPro: IPR000269; CUNH_oxidase.
DR Pfam: PF011179; Cu_amine-oxid; 1.
DR Pfam: PF02727; Cu_amine-oxid2; 1.
DR Pfam: PF02728; Cu_amine-oxid3; 1.
DR PRINTS: PR00766; CUDAOXIDSE.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
DR KEGG: OXIDOREDUCTASE; Copper; TPQ; Glycoprotein; signal; Metal-binding.
FT SIGNAL 1 16
FT CHAIN 17 762 COPPER AMINE OXIDASE, LIVER ISOZYME.
FT MOD_RES 470 470
FT METAL 519 519 COPPER (POTENTIAL).
FT METAL 521 521 COPPER (POTENTIAL).
FT METAL 683 683 COPPER (POTENTIAL).
FT BINDING 672 672 AMILORIDE (BY SIMILARITY).
FT CARBOHYD 136 136 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 231 231 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 665 665 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 762 AA; 84756 MW; AA959771360295FE CRC64;

Query Match 62.1%; Score 3326; DB 1; Length 762;
Best Local Similarity 81.8%; Pred. No. 37e-241;
Matches 606; Conservative 57; Mismatches 70; Indels 8; Gaps 1;

```



```

Db      319 GPRFVYQGNKVVASSLTFESFGCAESGPRVEDYRFQGERLAYETISLDAGAVTGGNTPPA 370
Oy      615 MTRRYVDGFGMKKYTPPLTRGVDCPYLATYVDHMFLLSESQAPKTINDACVEONQGLP 670
Db      379 MLTRYDMSGCMKYFATPTILRGVDCPYLATYMDMHFVESQTPTLTADCFVEONQGLP 430
Oy      675 LRRHSDLYSHYFGGLAEFLVLYVRKSNLTLDNYDWDTVFHPSCAIEIRFYANGYISSAF 730
Db      439 LRRHSDFLSHYEGGAQCVLFVFRSYSTMLNDYWDVWFYPNCAIEVKLHAICYISSAR 490
Oy      735 LFQGTGYGNGVSEHLTGTVHHSHAHEKVDLDVAGLEMMWAAEDMVFVPMAVPMSPPEHOL 790
Db      499 LEFGARRYGQNVGEHLTLGPVHHSAHYKKVDLDVGLEMMWAEDMAFVPALIPMSPEHOI 550
Oy      795 ORLOVTREKLLEMEEOAAFVSGATPRYTLYLASNSHNKMGCHGRYIOMLTSAGEPLRONS 850
Db      559 QRLDVTKOLETEEQAAFPFGCASPXYLYLASKOSKNKGHRGRIQYVSFASGPMQNOS 610
Oy      855 SMARGFSWERYQLAVTYQRKEEBSSSSVFNQNDMPAAPVDESDFINNETJAKDLVAWMT 910
Db      619 PMEAHFWMGRYQALAIORKETEPSSSSSVFNQNDPMTPTYDESDFINNETJAKDLVAWMT 670
Oy      915 AGFLHTIHADDINTYTYNGVGVCFLRPYNFEDEDPSRYSDASTYFRDDQAGACEVNPL 970
Db      679 AGFLHTIHADDIPTVYTVGVGVGFELRPYNFFDPQEPSMDSADSIYFREGDAGSCEINPL 730
Oy      975 ACLPQAACAAPDLPAFSHGCF 995
Db      739 ACLPQAATCAPDLPAFVSHGX 759

RESULT 5
AC AC2_HUMAN STANDARD; PRT: 756 AA.
ID AC02_106: 075105: 000120: 09UNU0:
AC 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Retina-specific copper amine oxidase precursor (EC 1.4.3.6) (RAO)
DE (Amine oxidase [copper-containing]).
GN AOC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OC NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Retina;
RX IMMURNA=97237047; PubMed=9119395;
RX IMMURA Y., Kubota R., Wang Y., Aakawa S., Kudoh J., Mashima Y.,
RX Oguchi Y., Shimizu N.;
RA "Human retina-specific amine oxidase (RAO): cDNA cloning, tissue
expression, and chromosomal mapping.",
RL Genomics 40:277-283(1997).
RN [2]
RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RX MEDLINE=98390194; Pubmed=9722954;
RA Imamura Y., Noda S., Mashima Y., Kudoh J., Oguchi Y., Shimizu N.;
RT "Human retina-specific amine oxidase: genomic structure of the gene
(RAO2), alternatively spliced variant, and mRNA expression in
retina.";
RL Genomics 51:293-298(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-Retina;
RA Zhang X., McIntire W.S.;
RT "Human copper-containing amine oxidases.",
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY BE A CRITICAL MODULATOR OF SIGNAL TRANSMISSION IN
RETINA, POSSIBLY BY DEGRADING THE BIOGENIC AMINES DOPAMINE,
HISTAMINE, AND PURINESINE.
CC -1- CATALYTIC ACTIVITY: RCH2(NH(2) + H(2)O + O(2) -> RCHO + NH(3)) +

```

CC	H(2)O(2).
CC	-1- COPFATOR: Binds 1 copper ion and 1 topaquinine per subunit.
CC	-1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
CC	HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC	-1- TISSUE SPECIFICITY: RETINAL-SPECIFIC.
CC	-1- PM: Topaquinine (TPO) is generated by copper-dependent
CC	autooxidation of a specific tyrosyl residue (by similarity),
CC	-1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; AB012943; BAA32590.1; .
DR	EMBL; AB012942; BAA32590.1; JOINED.
DR	EMBL; AB012943; BAA32589.1; .
DR	EMBL; AB012942; BAA32589.1; JOINED.
DR	EMBL; D88213; BAA19001.1; .
DR	EMBL; AF081363; AAD39345.1; .
DR	GeneW; HGNC:549; AOC2.
DR	MIM; 602268; .
DR	InterPro; IPR000269; CUNH_oxidase.
DR	Pfam; PF01179; Cu_amine_oxid. 1.
DR	Pfam; PF02727; Cu_amine_oxidn; 1.
DR	Pfam; PF02728; Cu_amine_oxidn; 1.
DR	PRINTS; PR00766; CU DAOXIDASE.
DR	PROSITE; PS01164; COPPER_AMINE_OXID.1; 1.
DR	PROSITE; PS01165; COPPER_AMINE_OXID.2; 1.
KV	Oxidoreductase: Copper; TPO; Alternative splicing; Glycoprotein;
KV	Signal; Metal-binding.
FT	SIGNAL 1 32
FT	CHAIN 33 756 POTENTIAL.
FT	MOD_RES 465 465 RETAINA-SPECIFIC COPPER AMINE OXIDASE.
FT	METAL 516 516 TOPAQUINONE (BY SIMILARITY).
FT	METAL 518 518 COPPER (POTENTIAL).
FT	METAL 680 680 COPPER (POTENTIAL).
FT	BINDING 669 669 COPPER (POTENTIAL).
FT	CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 226 226 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 588 588 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 662 662 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC 559 625 MISSING (IN SHORT ISOTORM).
FT	CONFLICT 181 181 D -> E (IN REF. 3).
FT	CONFLICT 215 218 RERT -> GDRA (IN REF. 3).
FT	CONFLICT 221 222 MA -> IG (IN REF. 1).
FT	CONFLICT 610 610 O -> H (IN REF. 3).
SEQ	SEQUENCE 756 AA: 83793 MW: 453B7317E0DACAB6 CRC64:

Query Match	47.7%	Score 2556	DB 1	Length 756
Best Local Similarity	65.8%	Pred. No. 1.8e-183		
Matches 480	Conservative	88	Mismatches 153	Indels 4
				Gaps 2
QY 268	GEPSOLPHCPSPVSPKOPTHPGOSOLFADLSHEETAAVRFRTORTLGGVDAQAARPS	327		
DB 29	GGSSOPPHCPSPVSHKQPPHPHPCQSOLFADLSHEETAAVRFRTORTLGGVDAQAARPS	88		
QY 328	DNCVSEVEIQLPPKAAALAHLDGRSPPARREALIATYFEGGQPOPNVSELVYGLPLPSPV	387		
DB 89	DNCIFSEVEIQLPPKAAALAHLDGRSPPARREALIATYFGGQPOPNVSELVYGLPLPSPV	148		
QY 388	RDVTYERHGGPLRYHRRPVLYFOEYLDIDIMIFENRFLPOASGLLHHCFFKHKRRGNLVYMT	447		
DB 149	RDVTYERHGGPLRYHRRPVLYRAEFTQWMRHLKVELPELPKPIFLST--FVNYGSTLAAYH	206		
QY 448	TAPRGISQDDPMTFELLYNISAGAEFLIHVGLLELVNKHADPARMTLOKVVYGGORYD	507		
DB 207	ADPRGSRSERTTMAALYHNISGVGLFRAEYVGLLELDHRAADPARMTYQOVYLYGHYYA	266		

Db 496 LQTHLLGIRHVLHYRVDMDVAGTKNSFQITMKLENTLNWSSHSIVOPTLQTOYS 555
QY 806 MEEQAFVGSATPRYLTLASNSKMGHPRGYRIQMLSFAGEPPLPONSSMARGFSEWERY 865
Db 556 QEHQAFVFGGTLPRKYLFLSSPOKNCWGRRRSRLQISMARQVLRPGQGEARVATMAYR 615
QY 866 QLVATQKEEPPSSSVENQNDNPAPYVDFSDFT-NNEFIACKDLVAVYTAGFLHHPAE 924
Db 616 PLATKYRESERYSSSLYNQNDPMDPVYFEFEFLRNNEDEDELVAVYTAGFLHHPSE 675
QY 925 DIPNVTGNGVGFRLRPVNEFDEPSPFSADSYFRGQDAGACEVNPACLPQAAACA 984
Db 676 DVPNATGNSYGFULRPFNEPDEPDSLASRDTIVY-WFQDGLNRVQ--RWIPEDRCL 732
QY 985 PDLPAFSHGC 994
Db 733 VS-PPEFSYNG 741

RESULT 7
ABP_HUMAN STANDARD: PRT: 751 AA.
ID AC P19801: Q16683: Q16684:
AD 01-FEB-1991 (Rel. 17, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Amiloride-sensitive amine oxidase [copper-containing] precursor
DE (EC 1.4.3.6) (Diamine oxidase) (DAO) (Amiloride-binding protein)
DE (ABP) (Histaminase).
GN ABPI OR DAO1 OR AOCL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE-94237856; PubMed-8182053;
RA Chassande O., Renard S., Barbry P., Lazdunski M.;
RT "The human gene for diamine oxidase, an amiloride binding protein.
RT Molecular cloning, sequencing, and characterization of the promoter.";
RL J. Biol. Chem. 269:14484-14489(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-Placenta;
RX MEDLINE-96113540; PubMed-8595053;
RA Zhang X., Kim J., McIntire W.S.;
RT "cDNA sequences of variant forms of human placenta diamine oxidase.";
RL Biochem. Genet. 33:261-268(1995).
RN [3]
RP PRELIMINARY SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-Kidney;
RX MEDLINE-91017502; PubMed-2217167;
RA Barbry P., Champe M., Chassande O., Munemitsu S., Champigny G.,
RA Lingueglia E., Maes P., Frelin C., Tartar A., Ullrich A.,
RA Lazdunski M.;
RT "Human kidney amiloride-binding protein: cDNA structure and functional
RT expression.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7347-7351(1990).
RN [4]
RP SEQUENCE OF 20-39, AND CHARACTERIZATION.
RC TISSUE-Placenta;
RX MEDLINE-94193685; PubMed-8144586;
RA Novotny W.F., Chassande O., Baker M., Lazdunski M., Barbry P.;
RT "Diamine oxidase is the amiloride-binding protein and is inhibited by
RT amiloride analogues.";
RL J. Biol. Chem. 269:9921-9925(1994).
CC -1- FUNCTION: CATALYZES THE DEGRADATION OF COMPOUNDS SUCH AS
CC PUTRESCINE, HISTAMINE, SPERMINE, AND SPERMIDINE. SUBSTRANCES
CC INVOLVED IN ALLERGIC AND IMMUNE RESPONSES, CELL PROLIFERATION,
CC TISSUE DIFFERENTIATION, TUMOR FORMATION, AND POSSIBLY APOPTOSIS.
CC PLACENTAL DAO IS THOUGHT TO PLAY A ROLE IN THE REGULATION OF THE
CC FEMALE REPRODUCTIVE FUNCTION.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +

CC H(2)O(2).
CC -1- COPACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: PLACENTA AND KIDNEY.
CC -1- PTM: Topaquinine (TPO) is generated by copper-dependent
CC autooxidation of a specific tyrosyl residue.
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
CC or send an email to license@sdb.ch).
CC
DR EMBL: X78212; CAA5046.1; -;
DR EMBL: U11862; AAC50270.1; -;
DR EMBL: U11863; AAB60381.1; -;
DR EMBL: M55602; AAA58358.1; ALT_SEQ.
DR PIR: A38276; A38276.
DR PIR: S42495; S42495.
DR Gene: HGNC:80; ABPI.
DR MIM: 104610; -;
DR InterPro: IPR00269; CUNH_Oxidase.
DR Pfam: PF01179; Cu_amine_oxid_1.
DR Pfam: PF02727; Cu_amine_oxidn2_1.
DR Pfam: PF02728; Cu_amine_oxidn3_1.
DR PRINTS: PR00766; CUPROXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
DR Signal: Glycoprotein; Oxidoreductase; Copper; Heparin-binding; TPO;
KW Alternative splicing; Polymorphism; Metal-binding.
FT SIGNAL 1 19
FT CHAIN 20 751
FT METAL 391 391
FT MOD_RES 461 461
FT METAL 510 510
FT METAL 512 512
FT METAL 675 675
FT BINDING 568 575
FT CARBOHYD 664 664
FT CARBOHYD 110 110
FT CARBOHYD 168 168
FT CARBOHYD 538 538
FT CARBOHYD 745 745
FT VARSPLIC 619 619
FT VARIANT 645 645
FT CONFLICT 28 28
FT CONFLICT 332 332
FT SEQUENCE 751 AA; 85363 MW; 1B8B6595C44232DA CRC64;
Query Match 26.0%; Score 1394; DB 1; Length 751;
Best Local Similarity 39.8%; Pred. No. 1, 9e-96;
Matches 293; Conservative 124; Mismatches 281; Indels 38; Gaps 11;
QY 285 PWTHPGOSOLFADLSRELVAVNRFLRGPGVLDAAQARPSPNCVSVELQPPKAA 344
Db 23 PGLTPKRAVGSFSDISNOELAAVSHFLMSKELRLOPSSITTMKNTYFLEMLPKRYH 82
QY 345 LAHLDKRGSPPEALALVFEFGROPVNSVGLVPLPHPSYMDVTVVERHGGPLPYHRR 404
Db 83 LRFLEDKGERPVREARAVIFFGDEHPNTEFAVGPLPGCYMRALS-PPRGYSSWASR 141
QY 405 PVLFOEYLDIDQMFNRELQASGLMH-----CCFYKRGKNTLVMTTAPGLOSQ 457
Db 142 PISRAEY-----ALVYHTLOEATKRPLOHFLNTTGFSDCHDCLAFVDAVGRVAGSQ 196
QY 458 RATWFGLYYNSAGFLHNVGLLELVNHRKALDPAWTIQKVYQGRYDLSLAOLEAOF 517

```

Db 197 RRSLLIQRVE--GYFLHPTGLTLLVDHGSTDAGHNAVEQVYNGKFKYSPEELARKYA 254
QY 518 AGLVNVLIPDNGTGGMSLSKSPVP-----GPAP-----PLQFYQGRFVSQ 562
Db 255 DGEVDVVLLEDPLEPGGSHDSTPEPLPSSHKRGDPSPSPHVSQPLVPGHGRFRLEG 314
QY 563 SRVASSLMTSFGLAFSPRIEDVAFQGERLYEISLQELALTYGNGSPAAMTTRVYDG 622
Db 315 NAVLYGMSFAFRSSSGQLVNLVHGERINAEVSGQEVNALKGHTPGMQTKLVDV 374
QY 623 GFGMGKTTPLTLCVDCPYLATYVDHMLLESQAPKTIRDAFCVEEQNOGLPLRRHNSDL 682
Db 375 GWSIGSTHLELPGIDCEPEATFELDTFHYDDADDPVHYPRALCLFEMPTGVLRRHNSN 434
QY 683 YS---HVEGLAEVLYVRESMTLLANDYWDYFHSGLAEIFYATGYTSAPFLGAT 739
Db 435 FKGGFNAYLKGQVLTAKRTSTVYNDYIWDPIFYNGVMEAKHATGYVHAFTYPEG 494
QY 740 GKYGNQVSEHTLTGVTHTSAHFVLDVAGLENNVMAEDVYFVMAVMPSEHLOLQV 799
Db 495 LRHGTRLHTLHNLGNHNLHYNVDLVAGTKNSFOLQKLENTITPMSFRHNVQPTL 554
QY 800 TRLLMEEOAALVGSATPRYLYLASNSHNGHPRGYRIQLSFAGEPLPQNSMARG 859
Db 555 EEOQYSEMERQAAFRFKRLPKYLLFTSPQENPWGHRKSYRLQHSMDQVLPQWQEQ 614
QY 860 FSEMYOLATORKKEEPPSSSVFNQNDPAPVYDSDFI--NNETIGKDLVAANTAGFL 918
Db 615 ITARPYPLATKYRESLCSSTIYHONDPMWHPVVEQFPHNNENINEDLVAVATVGL 674
QY 919 HIPHADIPVTYVNGVGVGFFLRPNFEFDEDPFSYADSITFYFGDDAGACEVPLACL 978
Db 675 HIPSSEDIPATATPGNSVGLLRPFNFEPEDPSLASMDYIV--WPRNGENYVO--RWIP 731
QY 979 QAACAPDLPAPSHG 994
Db 732 EDRDCSP--PPFSYNG 746

RESULT 8
AOC3_RAT
ID AOC3_RAT STANDARD: PRT; 321 AA.
AC 008590;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Membrane copper amine oxidase (EC 1.4.3.6) (Vascular adhesion protein-1) (VAP-1) (V997) (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=Sprague-Dawley; TISSUE=Adipocyte;
RC MEDLINE=97238878; PubMed=9083076;
RA Morris N.J., Ducret A., Aebersold R., Ross S.A., Keller S.R.,
RA Lienhard G.E.;
RT "Membrane amine oxidase cloning and identification as a major protein in the adipocyte plasma membrane."
RT J. Biol. Chem. 272:9388-9392(1997).
RL (2)
RN SEQUENCE OF 1-19.
RP TISSUE=Adipocyte;
RC MEDLINE=96010417; PubMed=8520629;
RA Jochen A., Guven S., Hays J.;
RA "The major integral membrane glycoprotein in adipocytes is a novel 200-kDa heterodimer."
RT Mol. Membr. Biol. 12:277-281(1995).
CC -1- FUNCTION: CELL ADHESION PROTEIN THAT PARTICIPATE IN LYMPHOCYTE RECIRCULATION BY MEDIATING THE BINDING OF LYMPHOCYTES TO PERIPHERAL LYMPH NODE VASCULAR ENDOTHELIAL CELLS IN AN L-SELECTIN-

```

```

CC INDEPENDENT FASHION. HAS A MONOMERIC OXIDASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) + H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
CC -1- PTM: Topaquinine (TPQ) is generated by copper-dependent autooxidation of a specific tyrosyl residue (by similarity).
CC -1- PTM: N- AND O-GLYCOSYLATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U72632; AAC5189.1; -.
DR InterPro; IPR000269; CuNH_oxidase.
DR Pfam; PF02727; Cu_amine_oxid2; 1.
DR Pfam; PF02728; Cu_amine_oxid3; 1.
DR PRINTS; PR00766; CUOAOXIDASE.
DR PROSITE; PS01164; COPPER_AMINE_OXID_1; PARTIAL.
DR PROSITE; PS01165; COPPER_AMINE_OXID_2; PARTIAL.
KW Oxidoreductase; Copper; TPQ; Glycoprotein; Transmembrane;
KW Signal-anchor; Cell adhesion; Metal-binding.
FT INIT-MET 0 0
FT DOMAIN 1 5
FT TRANSMEM 6 26
FT FT CYTOPLASMIC (POTENTIAL).
FT FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT (POTENTIAL).
FT FT EXTRACELLULAR (POTENTIAL).
FT CARBOHYD 136 >321
FT CARBOHYD 231 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON-TER 321 321
SQ SEQUENCE 321 AA; 35116 MW; 6B1F294253A1D6F CRC64;

Query Match 24.5%; Score 1315; DB 1; Length 321;
Best Local Similarity 82.7%; Pred. No. 4.5e-91;
Matches 244; Conservative 19; Mismatches 32; Indels 0; Gaps 0;

QY 263 QGDDGGPSQLPHCPSPSPAOQWTHPGSOLPADISREELAVNMFRLQRLPGVLDA 322
Db 27 RSDSDGRLSQLPHCPSPVLPSPQTHSGSQSPFADISPEELTAVMSFLNHLPLGLDAA 86
QY 323 QARPSPNCVSEVLEQPPKAAALAHLDGSPAPAREALATVFGRQPPVSELYGCLP 382
Db 87 QARPSPNCVSEVLEQPPKAAALAHLDGSPAPAREALATVFGRQPPVSELYGCLP 146
QY 383 HPSYMDVYVERHGGPLPYHRRPVEFOEYLDIDOMTFNRELPOASGLHHCFFYKGRN 442
Db 147 HPSYMDVYVERHGGPLPYHRRPVEFOEYLDIDOMTFNRELPOASGLHHCFFYKGRN 206
QY 443 LVYMTAPRGLSGDPAITWGLYNSGAGFLAHVGLLVNKKALDPAKMTIQKVFYO 502
Db 207 LKMTTAPRGLSGDPAITWGLYNSGAGFLAHVGLLVNKKALDPAKMTIQKVFYO 266
QY 503 GRYYDSLAOLEAOFEGAGLVNVLIPDNGTGGMSLSKSPVPGAPAPLQYPOQPR 557
Db 267 GRYYDSLAOLEAOFEGAGLVNVLIPDNGTGGMSLSKSPVPGAPAPLQYPOQPR 321

RESULT 9
GT26_SCHJA
ID GT26_SCHJA STANDARD: PRT; 218 AA.
AC P08515;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Glutathione S-transferase 26 kDa (EC 2.5.1.18) (GST 26) (Sj26 antigen) (GST class-alpha).

```

CC Schistosoma japonicum (Blood fluke).
 OC Eukaryota: Metazoa: Platyhelminthes: Trematoda: Digenea: Strigeidae;
 OC Schistosomatoidea; Schistosomatidae; Schistosoma.
 OX NCBI_TaxID=6182;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87041520; PubMed=3095841;
 RA Smith D.B., Davern K.M., Board P.G., Tiu W.U., Garcia E.G.,
 RA Mitchell G.F.;
 RT Weir 129/J mice is a parasite glutathione S-transferase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:8703-8707(1986).
 RN [2]
 RP REVISIONS.
 RA Smith D.B., Davern K.M., Board P.G., Tiu W.U., Garcia E.G.,
 RA Mitchell G.F.;
 RL Proc. Natl. Acad. Sci. U.S.A. 84:6541-6541(1987).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RX MEDLINE=95276631; PubMed=7538846;
 RA Lim K., Ho J.X., Keeling K., Gilliland G.L., Ji X., Ruker F.,
 RA Carter D.C.;
 RT "Three-dimensional structure of Schistosoma japonicum glutathione S-
 transferase fused with a six-amino acid conserved neutralizing
 RL epitope of gp41 from HIV."
 RL Protein Sci. 3:2233-2244(1994).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
 RX MEDLINE=95156484; PubMed=7853399;
 RA McTigue M.A., Williams D.R., Tainer J.A.;
 RT "Crystal structures of a schistosomal drug and vaccine target:
 RT glutathione S-transferase from Schistosoma japonica and its complex
 RL with the leading antischistosomal drug praziquantel.";
 RL J. Mol. Biol. 246:121-27(1995).
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
 CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
 CC HAEVATIN IN THE PARASITE GST.
 CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- MISCELLANEOUS: THERE ARE AT LEAST TWO ISOENZYMES OF GST IN
 CC S.JAPONICUM.
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC -----
 CC EMBL: M14654; AAB59203.1; -
 DR PIR: A26484; A26484.
 DR PDB: 1GNE; 30-NOV-94.
 DR PDB: 1GTA; 07-FEB-95.
 DR PDB: 1GTB; 01-DEC-95.
 DR PDB: 1B8X; 12-APR-99.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF000043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 KW Transferrase; Antigen; Multigene family; 3D-structure.
 SO SEQUENCE 218 AA; 25498 MW; 5E2AC418BD0EFL3F CRC64;
 Query Match 21.4%; Score 1144; DB 1; Length 218;
 Best Local Similarity 98.6%; Pred. No 1.6e-78;
 Matches 214; Conservative 0; Mismatches 3; Indels 0; Gaps 0

```

Db      2 SPLGKWKIKGLVQPRRLLEYLEEKYEHEHLVERDSGDKMRNKKFELGLFEPNLPYYIDG 61
QY      98 DVKLTQSMALIRYIADKHNNLGGSPKERAIEISMLEGAVLDIRGVSIATSKPEFELTKVD 157
Db      62 DVKLTQSMALIRYIADKHNNLGGCPKERAISMLEGAVLDIRGVSIATSKPEFELTKVD 121
QY      158 FLKSLPEMLKMFEDRLSHKTYTLNGDHTHDFMLYDGLDVLYMDPDCDAPFKLYSFK 217
Db      122 FLKSLPEMLKMFEDRLCHKTYTLNGDHTHDFMLYDGLDVLYMDPDCDAPFKLYCFKK 181
QY      218 RIEAIPQIDKYLKSSKRIYAMPLOGMQATFEGGDHPK 254
Db      182 RIEAIPQIDKYLKSSKRIYAMPLOGMQATFEGGDHPK 218

RESULT 10
GT27_SCHEMA
ID      GT27_SCHEMA          STANDARD;          PRT;          218 AA.
AC      P35661;
DT      01-JUN-1994 (Rel. 29, Created)
DT      01-JUN-1994 (Rel. 29, Last sequence update)
DT      01-NOV-1995 (Rel. 32, Last annotation update)
DE      Glutathione S-transferase 26 kDa (EC 2.5.1.18) (GST 26) (SM26/2
DE      antigen) (GST class-alpha).
OS      Schistosoma mansoni (Blood fluke).
OC      Eukaryota; Metazoa; Platyhelminthes; Trematode; Digenea; Strigeiida;
OC      Schistosomatoidea; Schistosomatidae; Schistosoma.
OX      NCBI_TaxID=6183;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE-92131046; PubMed-1775156;
RX      Wright M.D., Harrison R.A., Weider A.M., Newport G.R., Mitchell G.F.;
RT      "Another 26-kilodalton glutathione S-transferase of Schistosoma
RT      mansoni."
RL      Mol. Biochem. Parasitol. 49:177-179(1991).
CC      -1- FUNCTION: PARASITOL. OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC      OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC      -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC      PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
CC      SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
CC      HAEMATIN IN THE PARASITE GUT.
CC      -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC      -1- SUBUNIT: HOMODIMER.
CC      -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC      or send an email to license@sdb-sdb.ch).
CC      -----
DR      EMBL: M73624; -; NOT_ANNOTATED_CDS.
DR      PIR: A45556; A45556.
DR      HSSP: P08515; 1GTA.
DR      InterPro: IPR004046; GST_Cterm.
DR      InterPro: IPR004045; GST_Nterm.
DR      Pfam: PF00043; GST_C; 1.
DR      Pfam: PF02798; GST_N; 1.
KW      Transferase; Antigen; Multigene family.
SQ      SEQUENCE 218 AA; 25411 MW; D8D3EE902B36185 CRC64;

Query Match      18.4%; Score 985; DB 1; Length 218;
Best Local Similarity 82.9%; Pred. No. 1,3e-66;
Matches 180; Conservative 19; Mismatches 18; Indels 0; Gaps 0.

QY      38 SPLIGWKIKGLVQPRRLLEYLEEKYEHEHLVERDSGDKMRNKKFELGLFEPNLPYYIDG 97
Db      2 SPLIGWKIKGLVQPRRLLEYLEEKYEHEHLVERDSGDKMRNKKFELGLFEPNLPYYIDG 61
QY      98 DVKLTQSMALIRYIADKHNNLGGSPKERAIEISMLEGAVLDIRGVSIATSKPEFELTKVD 157
Db      62 DVKLTQSMALIRYIADKHNNLGGCPKERAISMLEGAVLDIRGVSIATSKPEFELTKVD 121
QY      158 FLKSLPEMLKMFEDRLSHKTYTLNGDHTHDFMLYDGLDVLYMDPDCDAPFKLYSFK 217
Db      122 FLKSLPEMLKMFEDRLCHKTYTLNGDHTHDFMLYDGLDVLYMDPDCDAPFKLYCFKK 181
QY      218 RIEAIPQIDKYLKSSKRIYAMPLOGMQATFEGGDHPK 254
Db      182 RIEAIPQIDKYLKSSKRIYAMPLOGMQATFEGGDHPK 218

QY      98 DVKLTQSMALIRYIADKHNNLGGSPKERAIEISMLEGAVLDIRGVSIATSKPEFELTKVD 157
Db      62 DVKLTQSMALIRYIADKHNNLGGCPKERAISMLEGAVLDIRGVSIATSKPEFELTKVD 121
QY      158 FLKSLPEMLKMFEDRLSHKTYTLNGDHTHDFMLYDGLDVLYMDPDCDAPFKLYSFK 217
Db      122 FLKSLPEMLKMFEDRLCHKTYTLNGDHTHDFMLYDGLDVLYMDPDCDAPFKLYCFKK 181
QY      218 RIEAIPQIDKYLKSSKRIYAMPLOGMQATFEGGDHPK 254
Db      182 RIEAIPQIDKYLKSSKRIYAMPLOGMQATFEGGDHPK 218

```


DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 KW Transferase; Antigen; Multigene family.
 FT INIT MET 0 0 BY SIMILARITY.
 FT CONFLICT 22 22 Y -> V (IN REF. 3).
 FT CONFLICT 110 111 DP -> VS (IN REF. 3).
 FT CONFLICT 189 189 A -> P (IN REF. 3).
 SQ SEQUENCE 220 AA: 25598 MW: 2789150B/5D10IF CRC64;
 Query Match 13.4%; Score 719.5; DB 1; Length 220;
 Best Local Similarity 59.2%; Pred. No. 1e-46;
 Matches 199; Conservative 34; Mismatches 54; Indels 1; Gaps 1;
 QY 41 LGYWKIKGLVQPTRLLEYLEEKYEHLXERDEGDKMKNKFEELGEPNLPYIDGVK 100
 DB 4 LGYWKIKGLVQPTRLLEYLEEKYEHLXERDEGDKMKNKFEELGEPNLPYIDGVK 62
 QY 101 LTQSMATIRYIADKHNMLGSPKRAEISMEGAVLDIRYGVSRVATSKOPELKVDFLS 160
 DB 63 LTQSLALIRYIADKHNMLGSPKRAEISMEGAVLDIRYGVSRVATSKOPELKVDFLS 122
 QY 161 KLPEMLKMFEDRLSHKTYLNDVHTHPDMLYDALDVLYLWDPCLDAFPKLVSEKRIE 220
 DB 123 DLPTTKMMSDFLGKKNPILKSTSVSHVDPMYELDAIRYELPHCLDHPMLQPMRSRIE 182
 QY 221 AIPQIDKYLKSSKIYAMPLOGMQATFGGDPHPPRSQSQ 258
 DB 183 ALPKIKAYMESNRFIKWPLNGWMAFGGDAPSPSEKK 220
 RESULT 13
 GT28_FASHE
 ID GT28_FASHE STANDARD; PRT: 217 AA.
 AC P31671;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Glutathione S-transferase 26 kDa 47 (EC 2.5.1.18) (GST47) (FH47)
 DE (GST class-alpha).
 OS Fasciola hepatica (liver fluke).
 OS Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 OC Echinostomida; Echinostomata; Fasciolidae; Fasciola.
 OX NCBI_TaxID=6192;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92155306; PubMed=1740183;
 RA Panaccio M., Wilson L.R., Cramer S.L., Wijffels G.L., Spithill T.W.;
 RT "Molecular characterization of cDNA sequences encoding glutathione S-
 transferases of Fasciola hepatica.";
 RL Exp. Parasitol. 74:232-237(1992).
 RN (2)
 RP ERRATUM.
 RX MEDLINE=94039664; PubMed=8224094;
 RA Panaccio M., Wilson L.R., Cramer S.L., Wijffels G.L., Spithill T.W.;
 RL Exp. Parasitol. 77:385-385(1993).
 RN (3)
 RP SEQUENCE OF 7-105 FROM N.A.
 RA Cramer S.L.;
 RL Patent number WO9008819, 09-AUG-1990.
 RN (4)
 RP X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).
 RX MEDLINE=98035725; PubMed=9367777;
 RA Rosjohn J., Fell S.C., Wilce M.C.J., Sexton J.L., Spithill T.W.,
 RA Parker M.W.;
 RT "Crystallization, structural determination and analysis of a novel
 parasite vaccine candidate: Fasciola hepatica glutathione
 S-transferase.";
 RL J. Mol. Biol. 273:857-872(1997).
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
 CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO

CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
 CC HAEMATIN IN THE PARASITE GUT.
 CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M77681; AAA29140.1; -;
 DR EMBL: A00996; CAA00121.1; -;
 DR PDB: 1FHE; 29-JUL-98.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 KW Transferase; Antigen; Multigene family; 3D-structure.
 FT INIT MET 0 0 T -> I (IN REF. 3) f
 FT CONFLICT 65 65 RIGF -> FEEL (IN REF. 3).
 FT CONFLICT 102 105 OFB8B8FE63029E03 CRC64;
 SQ SEQUENCE 217 AA: 25281 MW: 25281 MW: OFB8B8FE63029E03 CRC64;
 Query Match 13.2%; Score 708; DB 1; Length 217;
 Best Local Similarity 58.2%; Pred. No. 7.3e-46;
 Matches 124; Conservative 37; Mismatches 52; Indels 0; Gaps 0;
 QY 41 LGYWKIKGLVQPTRLLEYLEEKYEHLXERDEGDKMKNKFEELGEPNLPYIDGVK 100
 DB 4 LGYWKIKGLVQPTRLLEYLEEKYEHLXERDEGDKMKNKFEELGEPNLPYIDGVK 63
 QY 101 LTQSMATIRYIADKHNMLGSPKRAEISMEGAVLDIRYGVSRVATSKOPELKVDFLS 160
 DB 64 LTQSLALIRYIADKHNMLGSPKRAEISMEGAVLDIRYGVSRVATSKOPELKVDFLS 123
 QY 161 KLPEMLKMFEDRLSHKTYLNDVHTHPDMLYDALDVLYLWDPCLDAFPKLVSEKRIE 220
 DB 124 ELPTTKMMSDFLGKKNPILKSTSVSHVDPMYELDAIRYELPHCLDHPMLQPMRSRIE 183
 QY 221 AIPQIDKYLKSSKIYAMPLOGMQATFGGDPHPPRSQSQ 253
 DB 184 ALPKIKAYMESNRFIKWPLNGWMAFGGDAPSPSEKK 216
 RESULT 14
 GT28_FASHE
 ID GT28_FASHE STANDARD; PRT: 217 AA.
 AC P31671;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Glutathione S-transferase 26 kDa 47 (EC 2.5.1.18) (GST7) (FH7) (GST
 DE class-alpha).
 OS Fasciola hepatica (liver fluke).
 OS Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 OC Echinostomida; Echinostomata; Fasciolidae; Fasciola.
 OX NCBI_TaxID=6192;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92155306; PubMed=1740183;
 RA Panaccio M., Wilson L.R., Cramer S.L., Wijffels G.L., Spithill T.W.;
 RT "Molecular characterization of cDNA sequences encoding glutathione S-
 transferases of Fasciola hepatica.";
 RL Exp. Parasitol. 74:232-237(1992).
 RN (2)
 RP ERRATUM.
 RX MEDLINE=94039664; PubMed=8224094;
 RA Panaccio M., Wilson L.R., Cramer S.L., Wijffels G.L., Spithill T.W.;


```

RN Exp. Parasitol. 77:385-385(1993).
RA [3]
RP SEQUENCE OF 8-217 FROM N.A.
RA Cramerl S.;
RL Patent number WO9008819, 09-AUG-1990.
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
CC HAEMATIN IN THE PARASITE GUT.
CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL: M77680; AAA29139.1; -
DR EMBL: A00994; CAA00119.1; -
DR HSSP: P31670; 1FHE.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C_1.
DR Pfam: PF02798; GST_N_1.
DR Transferase; Antigen; Multigene family.
KM INIT_MET 0 BY SIMILARITY.
FT NDKRMKG -> MIGNRGM (IN REF. 3).
FT CONFLICT 188 196 IKEYMSER -> SRYSMA (IN REF. 3).
FT CONFLICT 188 196 IKEYMSER -> SRYSMA (IN REF. 3).
SQ SEQUENCE 217 AA; 25196 MW; 0099E1F59E49A49E CRC64;

Query Match 12.9%; Score 690; DB 1; Length 217;
Best Local Similarity 58.7%; Pred. No. 1.9e-44;
Matches 125; Conservative 30; Mismatches 58; Indels 0; Gaps 0;

OY 41 LGYWKIKGLVPTRLLELEYEEHLYERDEGDKMRKKFELGLEFNPPLYIDGVK 100
DB 4 LGYWKIRGLQGVRLLELEYEEHLYERDEGDKMRKKFELGLEFNPPLYIDGVK 63
OY 101 LTOSMAIIRYIADKHNMLGSPKRAEISMLGAVLDIRGVSRISAKDFELKVDLS 160
DB 64 LTOSVAIMRYIADKHNMLGSPKRAEISMLGAVLDIRGVSRISAKDFELKVDLS 123
OY 161 KLPEMLKMFEDRLSHKTYLNGDHTVHPDMLYDALDVLYMDPKCLDAFPKLVSEKRIE 220
DB 124 GLPTLLKMSDFLGRDROLYLGSSVSHVDFMYEALDCIRYLAPOCLDEFPKLEKRSRIE 183
OY 221 AIPQIDKYLKSSKYIAMPLOGMOAFTFGGDDHP 253
DB 184 DLPKIKAYMESEKRIKMPPLNSHWSFSGGDDAP 216

RESULT 15
GT26_FASHE STANDARD: PRT: 217 AA.
AC P30112;
DT 01-APR-1993 (Rel. 25, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutathione S-transferase 26 kDa 51 (EC 2.5.1.18) (GST51) (FH51)
DE (GST class-alpha).
OS Fasciola hepatica (Liver fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
OC Echinostomida; Echinostomata; Fascioloidae; Fasciolidae; Fasciola.
OX NCBI_taxid=6192;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92155306; Pubmed=1740183;

```

```

RA Panaccio M., Wilson L.R., Cramerl S.L., Wajffels G.L., Splithill T.W.;
RT "Molecular characterization of cDNA sequences encoding glutathione S-
RT transferases of Fasciola hepatica.";
RL Exp. Parasitol. 74:232-237(1992).
RN [2]
RP ERRATUM.
RX MEDLINE=94039664; Pubmed=8224094;
RA Panaccio M., Wilson L.R., Cramerl S.L., Wajffels G.L., Splithill T.W.;
RL Exp. Parasitol. 77:385-385(1993).
RN [3]
RP SEQUENCE OF 9-217 FROM N.A.
RX MEDLINE=93228188; Pubmed=7692383;
RA Moro A., Rodriguez-Molina J.R., Hallyer G.V.;
RT "Sequence analysis of a Fasciola hepatica glutathione S-transferase
RT cDNA clone.";
RL Am. J. Trop. Med. Hyg. 48:457-463(1993).
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
CC HAEMATIN IN THE PARASITE GUT.
CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL: M77682; AAA29141.1; -
DR HSSP: P31670; 1FHE.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C_1.
DR Pfam: PF02798; GST_N_1.
DR Transferase; Antigen; Multigene family.
KM INIT_MET 0 BY SIMILARITY.
FT NDKRMKG -> MIGNRGM (IN REF. 3).
FT CONFLICT 134 134 N -> D (IN REF. 3).
FT CONFLICT 146 146 P -> T (IN REF. 3).
SQ SEQUENCE 217 AA; 25242 MW; EFC88FE7454ECC26 CRC64;

Query Match 12.9%; Score 689; DB 1; Length 217;
Best Local Similarity 58.7%; Pred. No. 1.9e-44;
Matches 125; Conservative 33; Mismatches 55; Indels 0; Gaps 0;

OY 41 LGYWKIKGLVPTRLLELEYEEHLYERDEGDKMRKKFELGLEFNPPLYIDGVK 100
DB 4 LGYWKIRGLQGVRLLELEYEEHLYERDEGDKMRKKFELGLEFNPPLYIDGVK 63
OY 101 LTOSMAIIRYIADKHNMLGSPKRAEISMLGAVLDIRGVSRISAKDFELKVDLS 160
DB 64 LTOSVAIMRYIADKHNMLGSPKRAEISMLGAVLDIRGVSRISAKDFELKVDLS 123
OY 161 KLPEMLKMFEDRLSHKTYLNGDHTVHPDMLYDALDVLYMDPKCLDAFPKLVSEKRIE 220
DB 124 GLPTLLKMSDFLGRDROLYLGSSVSHVDFMYEALDCIRYLAPOCLDEFPKLEKRSRIE 183
OY 221 AIPQIDKYLKSSKYIAMPLOGMOAFTFGGDDHP 253
DB 184 DLPKIKAYMESEKRIKMPPLNSHWSFSGGDDAP 216

Search completed: May 20, 2003, 11:35:06
Job time : 28.557 secs

```

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 10:12:24 ; Search time 68.8588 Seconds

(without alignments)
2986.328 Million cell updates/sec

Title: US-10-081-408-20

Perfect score: 5357
Sequence: 1 MDWLRLFLMAAQSINAA.....QAAACAPDLPAFSGHGFSSN 998

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp-archaea:*
- 2: sp-bacteria:*
- 3: sp-fungi:*
- 4: sp-human:*
- 5: sp-invertebrate:*
- 6: sp-mammal:*
- 7: sp-mhc:*
- 8: sp-organelle:*
- 9: sp-phage:*
- 10: sp-plant:*
- 11: sp-rodent:*
- 12: sp-virus:*
- 13: sp-vertebrate:*
- 14: sp-unclassified:*
- 15: sp-virus:*
- 16: sp-bacteriap:*
- 17: sp-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3405	63.6	765	6	09TRK6
2	3370	62.9	765	11	09R055
3	1404.5	26.2	751	11	08VC36
4	1032	19.3	447	11	08R229
5	1029	19.2	218	5	094745
6	721	13.5	218	5	025595
7	707	13.2	218	5	09XYL9
8	649	12.1	787	3	096X16
9	615.5	11.5	271	11	0920K6
10	518.5	9.7	218	6	09NOV4
11	513.5	9.6	219	13	09Q0M9
12	507.5	9.5	218	11	08R516
13	503.5	9.4	223	5	097117
14	484.5	9.0	218	6	09TSM5
15	481.5	9.0	218	6	09TSM4
16	481.5	9.0	221	6	09BE80

17	479.5	9.0	218	11	091Y83
18	472.5	8.8	204	11	09CW16
19	468.5	8.7	218	11	09WU21
20	464.5	8.7	218	11	09DC88
21	463.5	8.7	218	11	09DD25
22	463.5	8.7	219	5	027653
23	462.5	8.6	219	5	016058
24	455.5	8.5	225	4	096HA3
25	449.5	8.4	225	6	09BEA9
26	443.5	8.3	220	4	08WME1
27	441.5	8.2	232	11	09D538
28	440.5	8.2	225	11	0921B2
29	439.5	8.2	195	4	095465
30	415.5	7.8	188	6	09M2B4
31	388	7.2	741	1	09C6V7
32	386.5	7.2	219	5	09U582
33	381	7.1	654	16	0989X2
34	372.5	7.0	668	10	048552
35	368.5	6.9	660	17	097XW1
36	358	6.7	64	4	09UEU7
37	357.5	6.7	756	10	09ST12
38	357	6.7	194	4	09HBM7
39	353.5	6.6	650	10	023349
40	353	6.6	649	10	08W1C1
41	350	6.5	181	4	08TC98
42	339	6.3	735	10	09SM88
43	338.5	6.3	1794	10	09S168
44	335.5	6.3	422	10	09C6W1
45	330.5	6.2	220	5	087TE4

ALIGNMENTS

RESULT 1

ID 09TRK6 PRELIMINARY; PRT; 763 AA.

AC 09TRK6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Semicarbazide-sensitive amine oxidase (EC 1.4.3.6).
GN SSO.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovine; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Iwabuchi H., Matsumura K., Mure M., Kuroda S., Tanizawa K.;
RT "Molecular cloning of semicarbazide-sensitive amine oxidase gene from
RT Bovine aorta.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB019242; BAA8896.1;
DR InterPro: IPR000269; CuniH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid.1.
DR Pfam: PF02727; Cu_amine_oxid.2; 1.
DR Pfam: PF02728; Cu_amine_oxid.3; 1.
DR PRINTS: PR00766; CUDAOXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; UNKNOWN_1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
KW Oxidoreductase.
SQ SEQUENCE 763 AA; 84500 MW; DB61ED9A89E71E90 CRC64;

Query Match 63.6%; Score 3405; DB 6; Length 763;

Best Local Similarity 85.1%; Pred. No. 9.5e-267;

Matches 626; Conservative 44; Mismatches 66; Indels 0; Gaps 0;

OY 263 OGGGGGPPSOLPHCPSPSPSAQPTHTPGOSOLFADLSREELTAVMRFTLQRLGGLVDAA 322
Db 28 RGGGGGAASOPHYCPSPSPVQPTHTPGQNLFPADLSREELTAVMRFTLQRLGGLVDAA 87

QY	323	QARSNDQVSVLEQLDPPKAAALAHDRGSPPARRELAIVFGROQOPVNSLTVGRLP	382
	88	QARSNDQVSVLEQLDPPKAAALAHDRGSPPARRELAIVFGROQOPVNSLTVGRLP	147
QY	383	HPSYMRDVTYERHAGCPPLPYHRAPLVFOEYLDIDOMINRELPOASGLHHCCFKHNGRN	442
Db	148	QPSIMRDVYTERHAGCPPLPYRRPVLREYLDIDOMITNRELPOAAGVLHHCCSKYGGGN	207
QY	443	LVTMTAPRGLSGDRATWFGLYYNNISGAFELHHVGLLELVNHNKALDPARMTIOKVFYO	502
Db	208	LVTMTAPRGLSGDRATWFGLYYNNISGAGYLPVGLLELVNHNKALDPARMTIOKVFYO	267
QY	503	GRTYDLSLAOLEAORFADLVNVLIPDNGTSGSNLSKPPVPGAPRLQTFPGOPRFSVOG	562
Db	268	GRTYDLSLAOLEAORFADLVNVLIPNNGTSGSNLSKQVPPGAPRLQTFPGOPRFSVOG	327
QY	563	SRVASSLMTSEFGGAGSCGRPIEDVPRQGERLVEISLOEALAIYGGNSPAAMTTRVDC	622
Db	328	SRVASSLMTSEFGGAGSGRPIEDIRQGERLAEISLOEALAIYGGNTPAAMLTRMDG	387
QY	623	GFGMGKTYTPLTRGVDCPYLATTVYDMHFLLESQAPKTIIDAPFCVFEONOGPLPRHNSDL	682
Db	388	CFGMGKFTATPLTRGVDCPYLATTVYDMHFLLESQAPRTLDHAFCVFEONKGLPRHNSDF	447
QY	683	YSHFPGGLAEVLVVRBSMSTLLANDVYDWDVYFHHSGAIEIRFATGVISSAFLEFGANGKY	742
Db	448	ISOYFGGVGVVLELVFRSVSTLLANDVYDWDVYFHHNGAIEVKFHAHTGVISSAFFEFGTKOKY	507
QY	743	GNQVSEHTLGTVHSHSAHKVDLDVAGLENNVAEDVAFVPMAPMSPEHOLQVYTRK	802
Db	508	GNQVRENTLGTVHSHSAHKVDLDVAGLENNVAEDVAFVPTTYPMSPPEHOLQVYTRK	567
QY	803	LLEMEQAAFLVGSATPRYILYLASNHSNKMGHPRGYRIOMLSFAGEBLPONSNMARGFSW	862
Db	568	QLETEQAAFLVGLGASPRYILYLASKOSNKMGHPRGYRIQVYTSFAGRLLPONSTERAIKW	627
QY	863	ERYQLATYQKKEEPPSSSVNQNDDPAPRYDESDFINNETIAGKDLVAVYTAGFLIHP	922
Db	628	GRYDLAVYQKKEEPPSSSVNQNDDPPTPYDFDFINNETIAGKDLVAVYTAGFLIHP	687
QY	923	AEDIPNTYVNGNGGFFLRAPNPFDEDEPSFSADSIYFRRGQDAGCEVNPILCTLPAAA	982
Db	688	AEDIPNTYVNGNGGFFLRAPNPFDEDEPSINSADSIYFOKHQDAGCEVNSLACLPRDPA	747
QY	983	CAPDLPAFASHGCGFSHN 998	
Db	748	CAPDLPAFASHGCGFTN 763	
RESULT 2			
O9R055			
AC	O9R055	PRELIMINARY;	PRT; 765 AA.
AC	O9R055		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)		
DE	Copper amine oxidase (EC 1.4.3.6).		
GN	ACO3.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murineae; Mus.		
OX	NCBI_TaxID=10090;		
OX	(1)		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE=99194802; PubMed=10092636;		
RA	Moldes M., Feve B., Paltrait J.;		
RT	"Molecular cloning of a major mRNA species in murine 3T3 adipocyte		
RT	lineage, differentiation-dependent expression, regulation, and		
RT	identification as semicarbazide-sensitive amine oxidase.";		
RL	J. Biol. Chem. 274:9515-9523(1999).		
CC	-1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +		
CC	H(2)O(2).		
CC	-1- COFACTOR: COPPER AND TOPAOUNONE (BY SIMILARITY).		

ID	Q8VC36	PRELIMINARY:	PRT:	751 AA.
CC	-1: SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.			
DR	EMBL; AF115411; AAD09199.1; -.			
DR	MGD; MG1:1306797; Aoc3.			
DR	InterPro; IPR000269; CuNH_oxidase.			
DR	pfam; PF011179; Cu_amine_oxid; 1.			
DR	pfam; PF02727; Cu_amine_oxidN; 1.			
DR	pfam; PF02728; Cu_amine_oxidN3; 1.			
DR	PRINTS; PR00766; CUDAOXIDASE.			
DR	PROSITE; PS01164; COPPER_AMINE_OXID_1; 1.			
DR	PROSITE; PS01165; COPPER_AMINE_OXID_2; 1.			
KW	Copper; Oxidoreductase; TPO.			
SW	SEQUENCE 765 AA: 84534 MW: 4252854759F316C7 CXC64;			
QY	Query Match	62.9%;	Score 3370;	DB 11; Length 765;
QY	Best Local Similarity	83.1%;	Pred. No. 6.5e-264;	
QY	Matches 611; Conservative 61; Mismatches 63; Indels 0; Gaps 0;			
QY	263	OGGDGGEPSQLPHCGSVSPSAOPMTHNPOSOQLFADLSREELTAVNRFLTORLPGELVDA	322	
QY	28	RSGDGGISQPLHCCSVLPSVQPRTHPSOSQFALSPBELTAVNSFLTKBLGPGELVDA	87	
QY	323	QARPDNCFVSVELDLPKKAALAHLDHSGSPREAREALATVEFGKQPOPNVSELVGGELP	382	
QY	88	QARPDNCFVSVELDLPKKAALAHLDHSGSPREAREALATVEFGKQPOPNVSELVGGELP	147	
QY	363	HPSYRDVTVERRHGRPLPYHRRPVLFQEYLDIDOMIFENBELFOASGLLHCCFYKRGKN	442	
QY	148	HPSYRDVTVERRHGRPLPYHRRPVLFQEYLDIDOMIFENBELFOASGLLHCCFYKRGKN	207	
QY	443	LYTMTTARPGLOSGRATVFGLYYISGAGFLVHVGLELLVNHKALDPARKTICKVYQ	502	
QY	208	LYTMTTARPGLOSGRATVFGLYYISGAGFLVHVGLELLVNHKALDPARKTICKVYQ	267	
QY	503	GRYYSLAOLEQFAGLVNVVLPDNGTSGSGMSLSPVPAPAPLPQYPOGPFPSYOG	562	
QY	266	GRYYSLAOLEQFAGLVNVVLPDNGTSGSGMSLSPVPAPAPLPQYPOGPFPSYOG	327	
QY	328	SGVSSSLMAFSPGLAFSGPRIFDRFOGERVAVEISYQEAIALYIGNSPASMSTCYVDG	387	
QY	563	SRVASSLMTFSFGLAFSGPRIFDRFOGERLVEISLQEAALYIGNSPAAATTRYVDG	622	
QY	328	SGVSSSLMAFSPGLAFSGPRIFDRFOGERLVEISLQEAALYIGNSPAAATTRYVDG	387	
QY	623	GGMGKYYTPLTRGYDCCPLATYYVNMHFLLESQAQKTRIDACVFEQONGPLRRHNSD	682	
QY	388	SGFICKYSPILRGVDCPYLATYYVNMHFLLESQAQKTRIDACVFEQONGPLRRHNSD	447	
QY	683	YSHYFGGLAEYLYVVRSMSTLNTDYVMDTVPHPSGAIEIRFYATGYISSAFLFGATGKY	742	
QY	448	YSHYFGGLAEYLYVVRSMSTLNTDYVMDTVPHPSGAIEIRFYATGYISSAFLFGATGKY	507	
QY	743	GNQVSEHTLGTVHTSHAEKVDLDVAGLENNVMAEDMVFVPAVWVSPBEHOLQRLQVTRK	802	
QY	508	GNQVSEHTLGTVHTSHAEKVDLDVAGLENNVMAEDMVFVPAVWVSPBEHOLQRLQVTRK	567	
QY	803	LLMEMEQAALVGSATPRLLYLAASHNSKMGIPRGRIOMLSFADEPLPONSMAKGSW	862	
QY	568	LLMEMEQAALVGSATPRLLYLAASHNSKMGIPRGRIOMLSFADEPLPONSMAKGSW	627	
QY	863	EKYTOALVORKEEPESSSVFNQNPMAPTYVDFSPFINNETAGKDLVAMVYAGFLHPIH	922	
QY	628	EKYTOALVORKEEPESSSVFNQNPMAPTYVDFSPFINNETAGKDLVAMVYAGFLHPIH	687	
QY	923	AADINNTVYVGVGFFLLRPYNEFDEDPSPFSYASDIYFRGDDAGACEVNPPLACIPQAAA	982	
QY	688	AADINNTVYVGVGFFLLRPYNEFDEDPSPFSYASDIYFRGDDAGACEVNPPLACIPQAAA	747	
QY	983	CAPDLLPAESHGFSH 997		
QY	748	CAPEIPAFESHGFSH 762		

DT 01-FEB-1997 (TREMBLrel. 02, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE 26kd glutathione S-transferase.
 OS Schistosoma mekongi.
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidae;
 OC Schistosomatidae; Schistosomatidae; Schistosoma.
 OX NCBI_TaxID=38744;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=THAILAND.
 RX MEDLINE=9622295; PubMed=9561610;
 RA Grams S.V., Grams R., Korge G., Vivanant V., Upatham S.;
 RT "Cloning and sequencing of the 26 kDa glutathione S-transferase gene
 of Schistosoma mekongi."
 RL Southeast Asian J. Trop. Med. Public Health 28:570-574(1997).
 DR EMBL; Y07663; CA66894.1; -.
 DR HSSP; P08515; 1GTA.
 DR InterPro; IPR004046; GST_Cterm.
 DR InterPro; IPR004045; GST_Nterm.
 DR Pfam; PF00043; GST_C; 1.
 DR Pfam; PF02798; GST_N; 1.
 KW Transferase.
 SQ SEQUENCE 218 AA; 25486 MW; D3080620B19DE23A CRC64;
 Query Match 19.2%; Score 1029; DB 5; Length 218;
 Best Local Similarity 88.9%; Pred. No. 3.9e-75;
 Matches 193; Conservative 11; Mismatches 13; Indels 0; Gaps 0;
 OY 38 SPILGYWKIKGLVOPTRLLLEYLEKEYEHLVERDEGDKWRKKFELGLEPFLNLYIDG 97
 DB 2 APLIGYWKIKGLVOPTRLLLEYLEGEYERLYERNEGDAMREKFLGLEPFLNLYIDG 61
 OY 98 DVKLQSAIIRYIADKHNMLGSGPKERAISMLGAVLDIRYGVSRAYSKDFFETLVKD 157
 DB 62 DVKLQSAIIRYIADKHNMLGSGPKERAISMLGAVSDIRSGVSRAYSNDFFETLVKD 121
 OY 158 FLSKLPEMLKMFEDRLSHKTYLNGDHVTHPPDMLYDALDVLYMDPCIDAFPKLVSEFK 217
 DB 122 FLNKLPEMLKMFEDRLCHKTYLNGDRTVHPDMLYDALDVLYMDKCLDAFPKLVSEFK 181
 OY 218 RIEAIPQIDKYLKSSKYIAMPLOGMOATFGGDDHPK 254
 DB 182 RIEAIPQIDKYLKSSKYIAMPLOGMOATFGGDDHPK 218
 RESULT 6
 ID 025595 PRELIMINARY; PRT; 218 AA.
 AC 025595;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative glutathione transferase.
 OS Clonorchis sinensis.
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 OC Opisthorchiida; Opisthorchiata; Opisthorchiidae; Opisthorchiidae;
 OC Clonorchis.
 OX NCBI_TaxID=79923;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hong S.-J., Lee D.-H.;
 RT "Cloning and over expression of 26 kDa glutathione S-transferase from
 Clonorchis sinensis."
 RL Thesis (1996), Parasitology, Chung-Ang University College of Medicine,
 Seoul.
 RP SEQUENCE FROM N.A.
 RA Hong S.-J.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; I47592; AAB46369.3; -.
 DR HSSP; P31670; 1FHE.
 DR InterPro; IPR004046; GST_Cterm.
 DR InterPro; IPR004045; GST_Nterm.

DR Pfam; PF00043; GST_C; 1.
 DR Pfam; PF02798; GST_N; 1.
 KW Transferase.
 SQ SEQUENCE 218 AA; 25038 MW; 7CB17C7B837A0B7C CRC64;
 Query Match 13.5%; Score 721; DB 5; Length 218;
 Best Local Similarity 59.0%; Pred. No. 3.5e-50;
 Matches 128; Conservative 38; Mismatches 51; Indels 0; Gaps 0;
 OY 38 SPILGYWKIKGLVOPTRLLLEYLEKEYEHLVERDEGDKWRKKFELGLEPFLNLYIDG 97
 DB 2 APLIGYWKIKGLVOPTRLLLEYVGSYEHSHYRCDEKQMDKHNGLGLEPFLNLYIDG 61
 OY 98 DVKLQSAIIRYIADKHNMLGSGPKERAISMLGAVLDIRYGVSRAYSKDFFETLVKD 157
 DB 62 NSFLQSAIIRYIADKHNMLGSGPKERAISMLGAVLDIRYGVSRAYSKDFFETLVKD 121
 OY 158 FLSKLPEMLKMFEDRLSHKTYLNGDHVTHPPDMLYDALDVLYMDPCIDAFPKLVSEFK 217
 DB 122 YLQQLPSTLRMSQFLGNNYSYLGSTPTPHLDPMFEALDVIRYLDPTSGVEAFPNLMQFIH 181
 OY 218 RIEAIPQIDKYLKSSKYIAMPLOGMOATFGGDDHPK 254
 DB 182 RIEAIPQIDKYLKSSKYIAMPLOGMOATFGGDDHPK 218
 RESULT 7
 ID 09XYL9 PRELIMINARY; PRT; 218 AA.
 AC 09XYL9;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Glutathione S-transferase.
 GN GST-1.
 OS Fasciola gigantica.
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 OC Echinostomidae; Echinostomata; Fasciolidae; Fasciolidae; Fasciola.
 OX NCBI_TaxID=46835;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=THAILAND.
 RA Grams S.V., Grams R., Sobhon P., Vivanant V., Upatham S.;
 RT "Molecular cloning of expressed antigens from Fasciola gigantica."
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF112567; AAD23997.1; -.
 DR HSSP; P31670; 1FHE.
 DR InterPro; IPR004046; GST_Cterm.
 DR InterPro; IPR004045; GST_Nterm.
 DR Pfam; PF00043; GST_C; 1.
 DR Pfam; PF02798; GST_N; 1.
 KW Transferase.
 SQ SEQUENCE 218 AA; 25337 MW; E9423D75C3F2EAF CRC64;
 Query Match 13.2%; Score 707; DB 5; Length 218;
 Best Local Similarity 58.7%; Pred. No. 4.7e-49;
 Matches 125; Conservative 36; Mismatches 52; Indels 0; Gaps 0;
 OY 41 LGYWKIKGLVOPTRLLLEYLEKEYEHLVERDEGDKWRKKFELGLEPFLNLYIDG 100
 DB 5 LGYWKIKGLVOPTRLLLEYLEGEYERLYERNEGDAMREKFLGLEPFLNLYIDG 64
 OY 101 LTQSAIIRYIADKHNMLGSGPKERAISMLGAVLDIRYGVSRAYSKDFFETLVKD 160
 DB 65 LTQSAIIRYIADKHNMLGSGPKERAISMLGAVLDIRYGVSRAYSKDFFETLVKD 124
 OY 161 KLPEMLKMFEDRLSHKTYLNGDHVTHPPDMLYDALDVLYMDPCIDAFPKLVSEFK 220
 DB 125 ELPTLTKMSDFLDROYLTGSSVSHVDPMYFALDRIYLAPOCLDFFPKLFEKSRIE 184
 OY 221 AIPQIDKYLKSSKYIAMPLOGMOATFGGDDHPK 253
 DB 185 DIPKIKAYMESEKTIKMPILNSWTASFGGDDAP 217

RESULT 8

096X16 PRELIMINARY; PRT; 787 AA.

ID 096X16
AC 096X16
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Lysyl oxidase.
GN AOC1.
OS Pichia pastoris (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxID=4922;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21131996; PubMed=11237259;
RA Kucha J.A., Dooley D.M.,
RT "Cloning, sequence analysis, and characterization of the 'lysyl
oxidase' from Pichia pastoris."
RL J. Inorg. Biochem. 83:193-204(2001).
DR EMBL: AF358434; AAK4976.1; -
DR InterPro: IPR000269; CuNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
SQ SEQUENCE 787 AA; 89650 MW; D779E8F2BF66745 CRC64;

Query Match 12.1%; Score 649; DB 3; Length 787;
Best local similarity 27.3%; Pred. No. 1.9e-43;

Matches 201; Conservative 114; Mismatches 320; Indels 102; Gaps 21;

OY 280 SPSAOWTHPGOSQLPADLSREELTVAMRFLTRLPGLVDAQAQAPSDNCVSVELQPL 339
DB 55 APTNNTMT-----SLAKEVQOELDLNHTY--NITEVTKAFQFSYVLMITLKP 103
OY 340 PKAALAHLDGRSPPREALALVFGROPQPVNSELVLPHPSPMRDVTERRGGPL 399
DB 104 NKTEALITLDEDDLPFRNARTVYTGEGEGEYFEELKVGPL---VSDETITE---PL 156
OY 400 PYHRRPVLFQELYLDIDQMTFNRELPOASGLLHCCEYKARGRLVYTTAPRGL----- 453
DB 157 SEFNTNGSKSLPREVGHLRIKSAKSS-----FLKNLNTTIRVDVLEGLIGVPE 208
OY 454 -----QSGRATVFGLYT-----NISGAFELHNVLELLVNHKALDPA 492
DB 209 DMGCHSAAPQLHPATGATYDGTCTINTENDAEMLVPTGFEFR-----FDMTGRDVS 261
OY 493 RMTIOKVFYGGRYDLSAOL-EAOFEAGLVNVVLI-P-DNGTSGMSLSKSPVPPGPR----- 545
DB 262 QMKMELEYNNKYTSSEELYEAKQKDEFTTLKRIDVN---LDMYIQORDSAPRINHL 318
OY 546 --APPLQFPQCPREFSVQGSRVASL--WTFSEGLGAFSGPRIFDVAFQGERLYEISLQ 601
DB 319 DRSPRLVEDEGRMAYDAGEEYFSSMDMGFYTSMRDGTISFDITFKSGRIYELSLQ 378
OY 602 EALATIGGNSPRAMATTYYVDGEGMGKYYTPRLTRGVDCPIATATVDMH-ELLEQAOKTI 660
DB 379 ELIAEYSGDDPFNOHTFYSDISYGVGNRFS-LVPGYDCPRTAGYFTTDEYDEFYKRTL 437
OY 661 RDAFCVEQNOGLPLRRHNSDLSHYFGLAETVLVRSMTLNTYVDMVTVHPSGAI 720
DB 438 --STCVENQEDYSLNHTGASYSAT---TONPLTNVRFISTIGNYDINFLYKFFLDGTL 492
OY 721 EIREYATGYISSAFLFGAT-GKYGNQVSEHTLGTVHTSAHFVLDVAGLENNVMAEDM 779
DB 493 EVSVRAAGYIOAGYWNPEYTSAPYGLKIHDLVLSGSFHDVLYNKYDLVDCGKKNASKVYM 552
OY 780 VFEPMVAVPSEHQLQVTRKLEME-----QAFLVGSATPRYLYLASHS 829
DB 553 KDVOVEYFMAFGTYVNTKQIAREVLEKEDFNGINWPPENGQILLESF-----EET 603
OY 830 NKMGHPRGRTOMLSFAGEPLPONS-SMARGFSMERYOLAVTORKEEPPSSSVFNOND 888

DB 604 NSFGNPRAYNIMPGGCGVHRIYKNSRGPEIQNMARSNLFLLTKHKDEELRSSTALNTNAL 663
OY 889 WAPYVDFSDFINNETIACKDLVAMYTACGLHHPHAEIDPNTVYGVNGVGFLLRPYNFDE 948
DB 664 YDPVNFNAFLDESDLEDSDIAWVNLGLHLHPNSNDLPNTLFTSAHASFMLTPRYFDS 723
OY 949 DPFYSADSIYFPGDD 965
DB 724 ENSRDTQOVFTTYDE 740

RESULT 9

0920K6 PRELIMINARY; PRT; 271 AA.

ID 0920K6
AC 0920K6
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Diamine oxidase (EC 1.4.3.6) (Fragment).
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriocognathi; Cavidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HARTLEY; TISSUE=SMALL INTESTINE;
RA Kitahara J., Kitahara N., Takemura M.,
RT "Cloning of a cDNA for guinea-pig diamine oxidase (histaminase)."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBD databases.
DR EMBL: AB073298; BAB70498.1; -
DR InterPro: IPR000269; CuNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; UNKNOWN_1.
KW Oxidoreductase.
FT NON_TER
SQ SEQUENCE 271 AA; 31323 MW; CEC46D64EAA135D CRC64;

Query Match 11.5%; Score 615.5; DB 11; Length 271;
Best local similarity 49.1%; Pred. No. 1.7e-41;

Matches 115; Conservative 31; Mismatches 87; Indels 1; Gaps 1;

OY 726 ATGYISSAFLFGATGKYGNQVSEHTLGTVHTSAHFVLDVAGLENNVMAEDMVYVPA 785
DB 1 ATGYVNAFTFYTPGQRTGRLHTLGNHTHLHYVVDVAGTKKNSFOTLQMKLENT 60
OY 786 VPMSPHQRLQVTRKLEMEQAAFLVGSATPRYLYLASNSNKKGHRGRTIQLSF 845
DB 61 NPMSPGHNLVQPTLKOSRYSQERQAAFRFGQTLPKYLLFTSREKRNWGHGRGYRLQISM 120
OY 846 AGEFLPNSSMARGFSMERYOLAVTORKEEPPSSSVFNONDPAVYVDSDFI-NNETI 904
DB 121 AEQVLPRAGDEPERAITARPLAVTKYRESELSSSIYONDPDPVYVOKFLHNEDI 180
OY 905 AGKDLVAMVTAGFLHPHAEIDPNTVYGVNGVGFLLRPYNFEDDPSPYSADSI 958
DB 181 ENEDLVAMVTYGFLLHPHSEDVPTATPAGNSVGLPLRPNFKDPGLASNDTV 234

RESULT 10

09NOV4 PRELIMINARY; PRT; 218 AA.

ID 09NOV4
AC 09NOV4
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Class mu glutathione S-transferase.
GN GSM.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]

RP SEQUENCE FROM N.A.
 RC TISSUE-LENS:
 RA Jimenez-Arensio J.V., Garland D.:
 RT "A lens glutathione S-transferase, class mu, with thiol-specific
 RT antioxidant activity."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF249588; AAF64308.1;
 DR HSSP: P03488; 1GTV.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR003081; GST_mu.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR PRINTS: PR01267; GSTRNSFRASEM.
 KW Transferase.
 SQ SEQUENCE 218 AA; 25635 MW; 3D02EA0F43C07B0A CRC64;

Query Match 9.7%; Score 518.5; DB 6; Length 218;
 Best Local Similarity 47.8%; Pred. No. 8.8e-34;
 Matches 98; Conservative 36; Mismatches 66; Indels 5; Gaps 1;
 Oy 40 ILGWKIKGLVQPTRLLEYLEEKEEHLRYRDEG----DKWNRKKEFLGLEFPNLPYI 94
 Db 4 ILGWDIRGLAHAIIRLLLEYTDYEEQYSGADPDYDRSQWNEKFKLGIDFPNLPYL 63
 Oy 95 IDGDVKLTQSMALIRYIADKHNMLGSGPKERAISMLEGAVLDIRYGSRIAYSKDFETL 154
 Db 64 IDGHNKLTQSMALIRYIARKHNLCGTEPEEMIRVDILENOYADVLAARICYSDFEKL 123
 Oy 155 KVDLSKLPENLKMEEDRLSHKTYLNGDHYTHPDPMYDALDVLYMDPMCLDAFPKLV 214
 Db 124 KPGELKEPEKIKLFSEFLGRPFAGDKLYVDPLVYDVMHRIPEPKLDLAFPNLKD 183
 Oy 215 FKRIEAIPOIDKYLKSSKIYAMP 239
 Db 184 FISREGLKISAYMKSSRFLEPL 208

RESULT 11
 090WN9 PRELIMINARY; PRT; 219 AA.
 AC 090WN9;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Glutathione S-transferase (EC 2.5.1.18).
 GN GSTM1.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodidae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA De Luca A., Favaloro B., Sacchetti P., Angelucci S., Di Ilio C.:
 RT "Molecular cloning, expression and site-directed mutagenesis of a
 RT liver Mu-class glutathione S-transferase from Xenopus laevis."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ116988; CAD01094.1;
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 KW Transferase.
 FT CHAIN 2 219 XLGSTM1.
 SO SEQUENCE 219 AA; 25359 MW; 5F7FF8AEDDA82DFD CRC64;

Query Match 9.6%; Score 513.5; DB 13; Length 219;
 Best Local Similarity 47.9%; Pred. No. 2.2e-33;
 Matches 102; Conservative 34; Mismatches 72; Indels 5; Gaps 1;
 Oy 40 ILGWKIKGLVQPTRLLEYLEEKEEHLRYRDEG----DKWNRKKEFLGLEFPNLPYI 94
 Db 4 ILGWDIRGLAHAIIRLLLEYTDYEEQYSGADPDYDRSQWNEKFKLGIDFPNLPYL 63

Db 4 ILGWDIRGLAHAIIRLLLEYTDYEEQYSGADPDYDRSQWNEKFKLGIDFPNLPYL 63
 Oy 95 IDGDVKLTQSMALIRYIADKHNMLGSGPKERAISMLEGAVLDIRYGSRIAYSKDFETL 154
 Db 64 IDGHNKLTQSMALIRYIARKHNLCGTEPEEMIRVDILENOYADVLAARICYSDFEKL 123
 Oy 155 KVDLSKLPENLKMEEDRLSHKTYLNGDHYTHPDPMYDALDVLYMDPMCLDAFPKLV 214
 Db 124 KPGELKEPEKIKLFSEFLGRPFAGDKLYVDPLVYDVMHRIPEPKLDLAFPNLKD 183
 Oy 215 FKRIEAIPOIDKYLKSSKIYAMP 239
 Db 184 FISREGLKISAYMKSSRFLEPL 208

RESULT 12
 08R516 PRELIMINARY; PRT; 218 AA.
 ID 08R516;
 AC 08R516;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Glutathione transferase GSTM7-7 (EC 2.5.1.18).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Guo J., Zimlak L., Zimlak P., Orchard J.L., Singh S.V.:
 RT "Cloning and expression of a novel murine glutathione
 RT transferase isoenzyme."
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF464943; AAL76248.1;
 KW Transferase.
 SQ SEQUENCE 218 AA; 25519 MW; 81F15DBB46118102 CRC64;

Query Match 9.5%; Score 507.5; DB 11; Length 218;
 Best Local Similarity 45.3%; Pred. No. 6.8e-33;
 Matches 96; Conservative 40; Mismatches 71; Indels 5; Gaps 1;
 Oy 41 LGWIKIKGLVQPTRLLEYLEEKEEHLRYRDEG----DKWNRKKEFLGLEFPNLPYI 95
 Db 5 LGWDIRGLAHAIIRLLLEYTSSSTEERKTYGADPDYDRSQWNEKFKLGIDFPNLPYL 64
 Oy 96 DGDVKLTQSMALIRYIADKHNMLGSGPKERAISMLEGAVLDIRYGSRIAYSKDFETL 155
 Db 65 DSHKLTQSMALIRYIARKHNLCGTEPEEKIRVDILENOYADVLAARICYSDFEKL 124
 Oy 156 VDFLSKLPENLKMEEDRLSHKTYLNGDHYTHPDPMYDALDVLYMDPMCLDAFPKLV 215
 Db 125 VEYLEQPLGMYKLFPSOFGORTWVGEEKITFVDLAYDIDLHLIFETPCIDAFPNLKD 184
 Oy 216 KKRIEAIPOIDKYLKSSKIYAMP 239
 Db 185 VAREVLRKISAYMKSSRFLEPL 208

RESULT 13
 097117 PRELIMINARY; PRT; 223 AA.
 ID 097117;
 AC 097117;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Glutathione S-transferase.
 OS Bosophilus microphilus (cattle tick).
 OC Eukaryota; Metazoa; Arthropoda; Cheilicerata; Arachnida; Acari;
 OC Parasitiformes; Ixodida; Ixodidae; Boophilus.
 OX NCBI_TaxID=6941;
 RN [1]
 RP SEQUENCE FROM N.A.

RA MEDLINE-99381232; PubMed-10451925; G.W., George J.E.;
He H., Chen A.C., Davey R.B., Ivie J.E.,
RT "Characterization and molecular cloning of a glutathione S-transferase
gene from the tick, Boophilus microplus (Acari: Ixodidae).";
RL Insect Biochem. Mol. Biol. 29:737-743(1999).
DR EMBL: AF077609; AAD15991.1; -
DR HSSP: P20136; 1GSU.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR003081; GST_mu.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C_1.
DR Pfam: PF02798; GST_N_1.
DR PRINTS: PRO1267; GSTNSFRASEM.
KW Transferase.
SQ SEQUENCE 223 AA; 25604 MW; F276ACE2D8259174 CRC64;

Query Match 9.4%; Score 503.5; DB 5; Length 223;
Best Local Similarity 47.0%; Pred. No. 1.5e-32;
Matches 101; Conservative 31; Mismatches 78; Indels 5; Gaps 1;

OY 38 SPILGYWKIKGLVPTLLELEYEEHLYERD-----EGDKRNKKFELGLEPPLP 92
DB 2 APLVGYMDIRGLAOPFIRLLAHADAKYDDKRYCGPPEDRSSWLNKTKLGLPPLP 61
OY 93 YIIDGVKLTOSMAIRIYADKHNMLGSGPKRAEISMLEGAVLDIRYGVSRIVASKDFE 152
DB 62 YIIDGVKLTOSMAIRIYADKHNMLGSGPKRAEISMLEGAVLDIRYGVSRIVASKDFE 121
OY 153 TLKVPDLTKPEMLKFEEDRLSHKTYLNGDHTVHPDFMLYDALDVLYMDPMLCLDFPKL 212
DB 122 KLGKDYTLKPLPASKAFSDYLGTHKRFAGDNLTLYVDFIAYEMLAQHILFAPDCLKPAPNL 181
OY 213 VSEFKRIEAIPOIDKYLKSSKYIAMPLOGWQATFG 247
DB 182 KAFVDRIEALPHVAAYLKSDCKIWPPLNGDMASFG 216

RESULT 14
OQTSMS PRELIMINARY; PRT; 218 AA.
AC OQTSMS;
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Glutathione S-transferase mu-class subunit M1 (BC 2.5.1.18).
GN GSTM1.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE-20330602; PubMed-10869451;
RA Wang C., Bammler T.K., Guo Y., Kelly E.J., Eaton D.L.;
RT "Mu-class GSTs are responsible for aflatoxin B(1)-8, 9-epoxide-
RT conjugating activity in the nonhuman primate macaca fascicularis
RT liver.";
RL Toxicol. Sci. 56:26-36(2000).
DR EMBL: AF200709; AAF08539.1; -
DR HSSP: P09488; 1GTU.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR003081; GST_mu.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C_1.
DR Pfam: PF02798; GST_N_1.
DR PRINTS: PRO1267; GSTNSFRASEM.
KW Transferase.
SQ SEQUENCE 218 AA; 25577 MW; EE5BAD80F60C95EB CRC64;

Query Match 9.0%; Score 484.5; DB 6; Length 218;
Best Local Similarity 43.9%; Pred. No. 5e-31;

Matches 93; Conservative 37; Mismatches 77; Indels 5; Gaps 1;
OY 41 LGYWKIKGLVPTLLELEYEEHLYERDEG-----DKRNKKFELGLEPPLPYI 95
DB 5 LGYWKIRGLAHARLLEETDSSYEKKYTMGDADPYDRSQWLNKFKGLDPPLPYLI 64
OY 96 DGDVKTOSMAIRIYADKHNMLGSGPKRAEISMLEGAVLDIRYGVSRIVASKDFETLK 155
DB 65 DGHKTIOSNALIRIYADKHNMLGSGPKRAEISMLEGAVLDIRYGVSRIVASKDFETLK 124
OY 156 VDFLSKLPKMLKFEEDRLSHKTYLNGDHTVHPDFMLYDALDVLYMDPMLCLDFPKLVSF 215
DB 125 PEYLEGPLPMQHFSQFLKRPMEVGDKITFVDLAYDVLDIRHIREPKCLDAFPMLKDF 184
OY 216 KRIEAIPOIDKYLKSSKYIAMPLOGWQATFG 247
DB 185 ISHFEGLERISAYMKSSRFLPKPLYYRVAWG 216

RESULT 15
OQTSMS PRELIMINARY; PRT; 218 AA.
AC OQTSMS;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Glutathione S-transferase mu-class subunit M2 (BC 2.5.1.18).
GN GSTM2.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE-20330602; PubMed-10869451;
RA Wang C., Bammler T.K., Guo Y., Kelly E.J., Eaton D.L.;
RT "Mu-class GSTs are responsible for aflatoxin B(1)-8, 9-epoxide-
RT conjugating activity in the nonhuman primate macaca fascicularis
RT liver.";
RL Toxicol. Sci. 56:26-36(2000).
DR EMBL: AF200710; AAF08540.1; -
DR HSSP: P28161; 1HNA.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR003081; GST_mu.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C_1.
DR Pfam: PF02798; GST_N_1.
DR PRINTS: PRO1267; GSTNSFRASEM.
KW Transferase.
SQ SEQUENCE 218 AA; 25708 MW; F2E509C3949F9051 CRC64;

Query Match 9.0%; Score 481.5; DB 6; Length 218;
Best Local Similarity 44.8%; Pred. No. 8.7e-31;
Matches 95; Conservative 35; Mismatches 77; Indels 5; Gaps 1;

OY 41 LGYWKIKGLVPTLLELEYEEHLYERDEG-----DKRNKKFELGLEPPLPYI 95
DB 5 LGYWKIRGLAHARLLEETDSSYEKKYTMGDADPYDRSQWLNKFKGLDPPLPYLI 64
OY 96 DGDVKTOSMAIRIYADKHNMLGSGPKRAEISMLEGAVLDIRYGVSRIVASKDFETLK 155
DB 65 DGHKTIOSNALIRIYADKHNMLGSGPKRAEISMLEGAVLDIRYGVSRIVASKDFETLK 124
OY 156 VDFLSKLPKMLKFEEDRLSHKTYLNGDHTVHPDFMLYDALDVLYMDPMLCLDFPKLVSF 215
DB 125 PEYLEGPLPMQHFSQFLKRPMEVGDKITFVDLAYDVLDIRHIREPKCLDAFPMLKDF 184
OY 216 KRIEAIPOIDKYLKSSKYIAMPLOGWQATFG 247
DB 185 ISFEGLERISAYMKSSRFLPKPLYYRVAWG 216

Search completed: May 20, 2003, 11:37:53
Job time : 72.8588 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:39:34 ; Search time 41 Seconds

(without alignments)
2418.011 Million cell updates/sec

Title: US-10-081-408-2_COPY_20_763

Perfect score: 4010
Sequence: 1 LVCVLVLVGRGGDGGEPGSLP.....QAAACAPDLPAFSGHGFSSHN 744

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

```
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4010	100.0	763	20	AAV03219
2	2489.5	62.1	729	23	AAU84261
3	457.5	11.4	248	23	ABP41516
4	329	8.2	712	23	ABB91435
5	324	8.1	687	23	ABB92504
6	302.5	7.5	759	23	ABB92082
7	246.5	6.1	300	23	ABB92910
8	214.5	5.3	670	17	AA94370
9	127	3.2	863	21	AA842952
10	125.5	3.1	460	23	ABB92909

11	115	2.9	2675	21	AA807564	Protein encoded by
12	109	2.7	435	14	AA832999	Rat choline kinase
13	106.5	2.7	446	22	ABG13153	Novel human diapo
14	106	2.6	617	23	AAE23407	Lysine oxidase pro
15	106	2.6	825	22	ABG67526	Drosophila melano
16	104.5	2.6	594	22	ABG07023	Novel human diapo
17	104	2.6	1250	22	AA65914	Novel human diapo
18	103	2.6	2224	17	AAW04254	Novel human diapo
19	103	2.6	2224	17	AAW04254	Novel human diapo
20	102.5	2.6	541	20	AAV49564	Human Factor V. H
21	102.5	2.6	541	20	AAV01647	Cytochrome P450 de
22	102	2.5	1194	22	AA692078	Cytosolic protein
23	101.5	2.5	587	22	AA652856	Escherichia coli p
24	101.5	2.5	736	22	ABB57329	Escherichia coli p
25	101	2.5	617	23	AAE23403	Lysine oxidase co
26	100.5	2.5	718	22	AAU54939	Protonbacterium
27	100.5	2.5	1253	23	AAU10648	Human L1CAM protei
28	100.5	2.5	1257	20	AAW74152	Human L1 cell adhe
29	100.5	2.5	1544	23	AAU97541	Human phospholiposi
30	100	2.5	617	22	AAU02204	Escherichia coli p
31	98.5	2.5	844	22	ABB53083	Human secreted pro
32	98.5	2.5	897	22	AAU93422	Human secreted pro
33	98.5	2.5	933	20	AAV26349	Human secreted pro
34	98.5	2.5	933	21	AAV73400	Human secreted pro
35	98.5	2.5	1152	21	AAV94920	Human secreted pro
36	98.5	2.5	1494	23	AAU78460	Novel human diapo
37	98	2.4	810	22	ABG04698	Novel human diapo
38	98	2.4	929	20	AAV22189	Trimeric murine C3
39	97.5	2.4	308	20	AAV30811	Human secreted pro
40	97.5	2.4	903	21	AA842926	Human ORF2590
41	97	2.4	545	23	ABB91611	Herbicidally activ
42	96.5	2.4	763	22	AAU41222	Herbicidally activ
43	96.5	2.4	816	23	ABB89783	Human polypeptide
44	96.5	2.4	1137	8	AAV71182	Sequence of herpes
45	96	2.4	1982	22	ABG16404	Novel human diapo

ALIGNMENTS

RESULT 1	AAV03219	standard: Protein; 763 AA.
XX	AAV03219;	
AC	21-JUN-1999	(first entry)
XX		
DT		
XX		
DE		Amino acid sequence of the vascular adhesion protein-1.
XX		
KW		Human; vascular adhesion protein-1; VAP-1; endothelial cell;
KW		Lymphocyte; inhibition; amine oxidase...
XX		
OS		Homo sapiens.
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 43	/note= "potential O-glycosylation site"
FT	Misc-difference 47	/note= "potential O-glycosylation site"
FT	Misc-difference 679	/note= "potential O-glycosylation site"
FT	Misc-difference 679	/note= "potential O-glycosylation site"
FT	Misc-difference 137	/note= "potential N-glycosylation site"
FT	Misc-difference 232	/note= "potential N-glycosylation site"
FT	Misc-difference 294	/note= "potential N-glycosylation site"
FT	Misc-difference 592	/note= "potential N-glycosylation site"
FT	Misc-difference 618	/note= "potential N-glycosylation site"
FT	Misc-difference 666	/note= "potential N-glycosylation site"

Query Match	Best Local Similarity	100.0%	Score 4010;	DB 20;	Length 763;
Matches 744;	Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;
1	LVCVLLVGGGSGGGERSQLPHCPSPVSPSQPPTTHPGQSLFADLSREELTAVMRFLTORL	60			
20	LVCVLLVGGGSGGGERSQLPHCPSPVSPSQPPTTHPGQSLFADLSREELTAVMRFLTORL	79			
61	GPGLVDAARARPSDNCVFSEVLQLPKKAALAHLDGSPPARREALAIVFGROPQPNYS	120			
80	GPGLVDAARARPSDNCVFSEVLQLPKKAALAHLDGSPPARREALAIVFGROPQPNYS	139			
121	ELVVGGLPSPSTWRDVTVERHGGLPLPYARRPVLFOEYLDIDDMIFNRELPPQASGLLHCC	180			

Db	140	ELVYGPLPHSPYMDYVERHGGGLPRHRRPVLFQELVLDIDQMFHNELPQAAGLLHNC	199
Qy	181	FYKHGRNLVTMTAPRGLOSGDRATVFGLYYNISGAFGLHHVGELELLVNHKALDPAM	240
Db	200	FYKHGRNLVTMTAPRGLOSGDRATVFGLYYNISGAFGLHHVGELELLVNHKALDPAM	259
Qy	241	TIQKVFYGRYDLSLAOLEAFEGVGLVNVYLIPNGGSGNSLKSPPRPPRAPPLQFPQ	300
Db	260	TIQKVFYGRYDLSLAOLEAFEGVGLVNVYLIPNGGSGNSLKSPPRPPRAPPLQFPQ	319
Qy	301	GPRFVQGSRAVSSIMTSPFLGAFSGPRIDVFOGEBRLYEISLOEALATIGGNSPAA	360
Db	320	GPRFVQGSRAVSSIMTSPFLGAFSGPRIDVFOGEBRLYEISLOEALATIGGNSPAA	379
Qy	361	MTTRVYDGGFGMGKXTTPELTRGVDCPYLATVYDMHFLLESQAPRTINDACVFEQNGLP	420
Db	380	MTTRVYDGGFGMGKXTTPELTRGVDCPYLATVYDMHFLLESQAPRTINDACVFEQNGLP	439
Qy	421	LRHHSDLSYHFGGLATVLYVMSKSTLLNVDYVMTQVPHPSAIEIRFYATGYISSAF	480
Db	440	LRHHSDLSYHFGGLATVLYVMSKSTLLNVDYVMTQVPHPSAIEIRFYATGYISSAF	499
Qy	481	LEGATGKYGNOVSEHTLTQVTHSAHFVDDVAGLENNVAAEDMVFPVPAVPMSPENOL	540
Db	500	LEGATGKYGNOVSEHTLTQVTHSAHFVDDVAGLENNVAAEDMVFPVPAVPMSPENOL	555
Qy	541	QRIQYTRKLLKMEQAAFLVGSATPRYLLYLAASHNSNKGHPRGYRIQMLSFAGEPLPONS	600
Db	560	QRIQYTRKLLKMEQAAFLVGSATPRYLLYLAASHNSNKGHPRGYRIQMLSFAGEPLPONS	619
Qy	601	SWARFSPWERIOLAVTOKKEEPPSSSYFNONDPAPRVDSFTINNETIAGKDLVAVVT	660
Db	620	SWARFSPWERIOLAVTOKKEEPPSSSYFNONDPAPRVDSFTINNETIAGKDLVAVVT	679
Qy	661	AEFLIHIPHAEDIPNVTVTGNNGVGFRLRPYNPFDEDPSTYSADSITYFGDDODAGACEVNPPL	720
Db	680	AEFLIHIPHAEDIPNVTVTGNNGVGFRLRPYNPFDEDPSTYSADSITYFGDDODAGACEVNPPL	735
Qy	721	ACLPQAAACAPDLPAFSGHGSFN 744	
Db	740	ACLPQAAACAPDLPAFSGHGSFN 763	
RESULT 2			
AA084261			
ID	AA084261 standard; Protein: 729 AA.		
XX	AA084261;		
AC			
DT	08-MAY-2002 (first entry)		
XX			
DE	Human endometrial cancer related protein, AOC2.		
XX			
KW	Human; endometrial cancer; differential expression;		
XX	DNA microarray; protein microarray.		
OS	Homo sapiens.		
XX			
PN	WO200209573-A2.		
XX			
PD	07-FEB-2002.		
XX			
PF	31-JUL-2001; 2001WO-US24104.		
XX			
PR	31-JUL-2000; 2000US-221735P.		
XX			
PA	(BGHM) BRIGHAM & WOMENS HOSPITAL INC.		
XX			
PI	Mutler GL;		
XX			
DR	WPI: 2002-179967/23.		
XX			
DR	N-PSDB: ABR35481.		
XX			

PT Diagnosing endometrial cancer comprises determining expression of
PT nucleic acid molecules or expression products that are differentially
PT expressed in normal and malignant endometrium -
PS Claim 33; Page 131-134; 233pp; English.
XX
CC The invention relates to diagnosing endometrial cancer in a subject
CC suspected of having endometrial cancer comprising determining the
CC expression of a set of nucleic acid molecules or expression products in
CC an endometrial sample suspected of being cancerous, where the set of
CC nucleic acid molecules comprises at least 2 nucleic acid molecules
CC selected from 50 fully defined sequences as given in the specification.
CC The nucleic acids are used as an array of at least 2 of the 50
CC nucleic acids bound to a solid substrate. Also included is a solid-phase
CC protein microarray comprising at least 2 antibodies or its antigen
CC binding fragments, that specifically bind at least 2 different
CC polypeptides from the 50 fully defined sequences as given in the
CC specification, fixed to a solid substrate. The methods and arrays are
CC useful for the diagnosis of endometrial cancer, selecting and monitoring
CC treatment regimes and identification of lead compounds useful for the
CC treatment of endometrial cancer. The present sequence is one of 50
CC proteins differentially expressed between cancerous and non-cancerous
CC samples.
CC
CC
SQ Sequence 729 AA:
Query Match 62.1%; Score 2489.5; DB 23; Length 729;
Best Local Similarity 63.7%; Pred. No. 2,1e-244;
Matches 473; Conservative 86; Mismatches 149; Indels 35; Gaps 4;
OY 1 LVCVLVGGGCGGEPSPHPCPSVSPSAPWTHPCQSOLFADLSREELTAVKREFTOL 60
DB 20 LAYVLLTSPGS---SQPHCPSPVSHRAQPMHPHPCQSOLFADLSREELTAVKREFTOL 75
OY 61 GGLVDAQAARPSDNCVFSEVLOLPPKAAALAHLDGRGSPPPAREALATVFFGQPOPNTS 120
DB 76 GGLVDAQAARPSDNCVFSEVLOLPPKAAALAHLDGRGSPPPAREALATVFFGQPOPNTS 135
OY 121 ELVGPRLPHSYMDVTVRHHGGLPYHRRPVLFOEYLDIDOMIFNRELPOASGLLHCC 180
DB 136 ELVGPRLPHSYMDVTVRHHGGLPYHRRPVLFOEYLDIDOMIFNRELPOASGLLHCC 194
OY 181 FKHRRNRLVMTTAPRGOSGDRAWPLTYNISGAGFELHNVGLLVNKHALLPARK 240
DB 195 -FNYSNGSTLAHAHPRGARSRRRTWIGLHNISGVGLFHLRVGLLELDHRAALPAHM 253
OY 241 TIOKVEYQGRYYDSLQLEAOFEGALVNVLLIPDNGTGSWSLSPVPGPAPPLQFYPO 300
DB 254 TYQVQVYLLGHYYADLQGLREFRSGLRVYVPLPPNASSLRKNSFGPLPLQFSPQ 313
OY 301 GPRFSVOGRVASSLWTFESFGAIFSGLPIFDVRFQGERLVYEISLQALATYGNSPA 360
DB 314 GSOYVOGNLVVSSLSMFSFGHGVFSGRLIFDVRFOGERIAYEVSOECVSIXGADSPKT 373
OY 361 MTRRYVDGEGGKTYTPTLRGVDCPYLATVYDMHFLLESQAOKTIRDAFCVEQNGPL 420
DB 374 MLTRVLDSSFGGLGRNRSGLRVGDCPYQATMVDIHLVNGAQLLPGACVCEEAQGL 433
OY 421 LRRHSHDLTSHYFGALETVLVVRSMSTLNDYVDVTFHPSGAIEIRPATGYSSAF 480
DB 434 LRRHHYLDLNNHFFGGLASSALVYRSVSVGNVYIMDPVLPRGALGEGRAHATGYINTAF 493
OY 481 LFGATG--KYGNQVSEHTLGVTHSHAFKVDLDVAGLENMVAEDMVFVPAWPSPH 538
DB 494 LKGGEGLLFGNVRGVRVLTGVTHTAHFHEKLDVAGLKKMVAEADVFXPAVAPNPH 553
OY 539 QLORLQVTRKLLMEQAAFLVGSATPRYLYLASNSNKGHRGRIGRLQSLRAGEPLQ 598
DB 554 WLQRPOLQVLEKEDLTASLSSPLRYLYLASNOTNAGHQG----- 598
OY 599 NSSMAGFSEWERYQVLTQKREEPSSSSVFNQNDMPAPTVDSEFINNETIAGKDLVAN 658
DB 599 -----YQLVYTKQKEEEOSSSIYHONDITMTPIVTRADFINNETILGLEDLVAN 646

OY 659 VTAFGLHHPHADIDNTVYGVNGVGFLLRPYNFEDEDPSEYGSADSIYRGDDAGACEVN 718
DB 647 VTASFLLHHPHADIDNTVYLVGNRVGFLLRPNYFEDDEDPSEYGSADSIYRGDDAGACEVN 706
OY 719 PLACIPQAAACAPDLPAPSHGCF 741
DB 707 PVACLPDLAACVPLDPFSPYHGF 729
RESULT 3
ID ABP41516 standard; Protein; 248 AA.
XX
AC ABP41516;
XX
DT 22-AUG-2002 (first entry)
XX
DE Human ovarian antigen HVBK73, SEQ ID NO:2648.
XX
KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antiinflammatory; gynaecological; reproductive; chromosome 7q34-ter.
XX
OS Homo sapiens.
XX
PN WO200200677-A1.
XX
PD 03-JAN-2002.
XX
PF 07-JUN-2001; 2001WO-US18569.
XX
PT 07-JUN-2000; 2000US-209467P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Blize CE, Rosen CA;
XX
DR WPI: 2002-147878/19.
XX
DR N-PSDB; ABQ54593.
XX
PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g.
PT ovarian cancer), immune disorders, cardiovascular disorders and
PT neurological diseases -
PS
PS Claim 11; SEQ ID NO 2648; 2922pp; English.
CC
CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders

CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents a human ovarian antigen of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://wipo.int/pub/published_pct_sequences.

XX Sequence 248 AA;

Query Match 11.4%; Score 457.5; DB 23; Length 248;
Best Local Similarity 41.9%; Pred. No. 1.1e-37;
Matches 83; Conservative 45; Mismatches 67; Indels 3; Gaps 1;

QY 286 PVPPGAPPLQFYPOGPFVSGSRVASSLMTESFGAIFSPPRIFVRFQGERLYEIS 345
DB 16 PSPHVSGRPLVOPHGRFLEGNAVLVGMSFAFLRSSGLQVLNVHFGERIAYEVS 75
QY 346 LQELALTYGNSPRAAMTRIVDGGFGMGKTTPLTRGVDCPYLATYVDHMFLESQAPKT 405
DB 76 VOEVALYGGHTPAGMQTKYLDVGMGLSVTHELAPGIDCEPETAFLDTEHYVDADDPVH 135
QY 406 IRDAFCVFEQNOGLPLRRHSDLYS---HYFGLAETVLVVRSMSTLLNDYVMDVFEHP 462
DB 136 YPRALCFEHPYTPVPLRRHNSNMGKGFNFYAGLKGQVLVLTSTVYNDYIMDFIYF 195
QY 463 SGAIEIRFYATGYISSAF 480
DB 196 NGVWEAKMHATGYVHAXF 213

RESULT 4

ID ABB91435 standard; Protein; 712 AA.

XX ABB91435;

XX 31-MAY-2002 (first entry)

XX Herbicidally active polypeptide SEQ ID NO 646.

XX Herbicidal; plant; agriculture; herbicide.

XX Arabidopsis thaliana.

XX NC0200210210-A2.

XX 07-FEB-2002.

XX 28-AUG-2001; 2001WO-EP09892.

XX 28-AUG-2001; 2001WO-EP09892.

XX (FARB) BAYER AG.

XX Tietjen K, Weidner M.

XX WPI; 2002-269010/31.

XX Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plant with nucleic acid or amino acid sequences from non-plant
XX organisms -

XX Claim 5; SEQ ID NO 646; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins
XX (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX aligning and comparing nucleic acid or amino acid sequences from plant

CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.

XX Sequence 712 AA;

Query Match 8.2%; Score 329; DB 23; Length 712;
Best Local Similarity 22.1%; Pred. No. 8.2e-24;
Matches 160; Conservative 99; Mismatches 321; Indels 144; Gaps 25;

QY 44 LSREELTAVNRFLQRIQPGVDAQAARPSDNCFSVELQLPRAALAHLDRCSPPAR 103
DB 65 LTVREINRVRTILSNH-DPGSGSGAT-----HSHALDEPE-SRVQMKGNKLLSR 116
QY 104 EALVAIFEGROPNPVSELVGPRLPHPSYMDYVEHNGPLPRHRRPVLFQETLIDDM 163
DB 117 RAAVAVAMGQTHEITVDLSGRV---VSDYINRISGVPI-LTLNDVFAASGVPLKSL 170
QY 164 IFNRELQASGLHHCPCFYKRGKRLVMTTAPRGDSGRAT-----WGLYTNISGA 217
DB 171 EFNRSI-----EARGVFSDLACTTPRAGWFG----- 197
QY 218 GFELHVGLELLVNHKALDPAWRTQKVFYQGRYDSLAQEAQ---EAGLVNVVLPD 274
DB 198 ---SEEGRRYV-----RVQCFITLQTTNVRPLRLEGITYVVDLCKEIVKIID 243
QY 275 NGTGSWSLSPVPPGAPPLQF-----YPOGPFVSQV-GSRVASSL 315
DB 244 KG-----PIPIKASGTEFRFQVQKPVHMDRINPISMEQPGPSFRVEDGLVWMAN 296
QY 316 WTESFGAIFSPPRIFVRFQGERLYEISLQELALTYGNSPRAAMTRIVD----- 367
DB 297 WVFHVKADORAGMIY-----SQATVRDSETEGEPRSVYKGFSELFPYMDPEEGWYK 350
QY 368 -----GGFGMGKTYTPTLRGVDCPYLATYVDHMFLESQAPKTIRDAFCVFEQNOGLPLR 422
DB 351 GYNDAGELGLGPTAMPPLVPLNDCEPRNSYIDGVFASPDGKFIYQPPNACLEPRTAGDISW 410
QY 423 RHNSDLYSHYFGLAET---VLVVRSMSTLLNDYVMDVFEHPGAIEIRFYATG--- 474
DB 411 RHSEILFAN-ADIRSRPKVTIVARMAATSVGNDYIFDMEFQDGLIRTVAAASGLMV 468
QY 475 ----YISSAFLFGATGKYGNQVSEHTIGTYHTSHAHKVDLDVAGLENMTVAEDMVEVPM 530
DB 469 KGTPYDNDVDDLQREDDEAGPLISENVIGVHDHFTFHLDMDDIDGPMN---NSLVKVL 524
QY 531 ---AVPWSPEHQLORLQVTRKLEMEQAAFLVGSATPRYLTLAS-RHSNMKGHPRGYRI 586
DB 525 EKQRPVPGKSPKRSKYLVKKYIAKTEKDAQIKLSLYDPEFHYVNPKRKSRYGNPAGYRI 584
QY 587 QMLSFAGEPLPONSMAKGSWEERYQLAVYQKKEEPPSSSVFNQNDPMAPTVD-FSDFI 645
DB 585 VPGGNAASLLDHDPPQIRGAFTNQIWMVPRYNSSEQYAGGVLYQSGDDTLQWMSD-- 642
QY 646 NNERIAGKDLVAMWTAGFLHPRHEDIPNVTYVANGVGFLLRPNFDEDEPSFSADSIY 705
DB 643 RDRSIEKKDVLWTLTGFIHVPCEQDYVPMPYV--AASFELKPAHFESNPLIGSAD--F 698
QY 706 FRGD 709
DB 699 FEKD 702

RESULT 5

ID ABB92504 standard; Protein; 687 AA.

XX ABB92504;

XX 31-MAY-2002 (first entry)

```

XX Herbicidally active polypeptide SEQ ID NO 1715.
DE
DE Herbicidal; plant; agriculture; herbicide.
XX
XX Arabidopsis thaliana.
OS
XX WO200210210-A2.
XX
XX 07-FEB-2002.
XX
XX 28-AUG-2001; 2001WO-EP09892.
XX
XX 28-AUG-2001; 2001WO-EP09892.
XX
XX (FARB ) BAYER AG.
XX
XX Tietjen K, Weidler M;
XX
XX WPI: 2002-269010/31.
XX
XX Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plant with nucleic acid or amino acid sequences from non-plant
XX organisms -
XX
XX Claim 5; SEQ ID NO 1715; 261pp + Sequence Listing; English.
XX
XX The invention relates to identifying target proteins
XX (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX aligning and comparing nucleic acid or amino acid sequences from plant
XX with nucleic acid or amino acid sequences from non-plant organisms using
XX suitable search parameters, where plant sequences having an E-value
XX greater by a factor of 3 than the E-value of most similar non-plant
XX sequences are selected. The polypeptides or nucleic acids encoding them
XX are useful for identifying modulators. The identified modulators are
XX CC
XX
XX Sequence 687 AA:
SQ
Query Match 8.1%; Score 324; DB 23; Length 687;
Best Local Similarity 22.1%; Pred. No. 2.5e-23;
Matches 168; Conservative 102; Mismatches 298; Indels 192; Gaps 31;
OY 20 PHCPSPSPSAQPTWHPGQSLFADLSREELTAVMRFITQRLGCLVDAQAARPSDNCVFS 79
DB 39 PHRP-----LDPLTTP-----EIKRVOTILSGH-DPGFGSGS-----TIIHA 74
OY 80 VELQLPPKAAALAHLDGSPPARREALATVFFGRQPOPNV---SELVVGFLPHPSY---- 132
DB 75 MALDEPDKQKRVIMWKGDRLPRRAETLANSNESHVLTDLKSGRVASDVLNPTFGYPT 134
OY 133 --NRDVTVERHGRLPYHRHPRVFOEYLIDIDQMFNRELPOASGLHHCIFYKHGRNLY 190
DB 135 LTKMKDIAV---SOVPR-----KSEFNRSLSEARGIF-FSGILITTFPAGHYGPD-- 180
OY 191 TMTTAPRGLOSGRATWFGLYYNISGAGFLAHV-GLELLVNHKALDPAKMTIOKVEYOG 249
DB 181 -----EGRRRVIRKICFSSKODTVNFYMRPIEGLYLVDMDKTE----- 218
OY 250 RYDLSLAQLAEAFAGLVNVLVLPDNGTGSWSLKSPPVPGAPRPLDF----- 297
DB 219 -----TIKIVDNG-----PVVPVPSSTGTETRYVYGFLETVYMDRV 252
OY 298 -----YPOGPRFSVQ-GSRVASSLMTFSFGIGAFSGPRJEDVAFQGERLYVEISLQDAL 350
DB 253 NPMKMEPDGSPSFOVEGILYKANKMKFHFKPDORAGMIT-----SQATVRDCKTGEAR 306
OY 351 AI-YGNSPAA-----TTRYVDGQ-FGMGKYTTPPLRGVDCPYLATYVDMHFL 397
DB 307 SVMKKGFASELFVYNNMDPGEGWYSKAYMDAGEFGLGSSMPLVLDNCPRNAAYIDGFFA 366
OY 398 LESQAPRTINDAPCVFQDNGCLPLRRHHSDLXHYFGGL-----AETVLVYVRMSSTLLN 451

```

```

DB 367 SPEGIPIDLPNMICLFERRAGDTSWRHSEIL-----LPQVDINESRAKATLVARMACSVGN 422
OY 452 YDYVMDTVFHPGSAIEIRFYATG-----YISSAFLEFGATGKGNQVSEHTGLTVHMH 503
DB 423 YDIYFDMFQMDGIVRTVAASGMLWKGTAVERNEDLEKEKDDSGPLISENVIQVYHGH 482
OY 504 SAHFKYDLVAGLENNVNAEDMVFVPMVPMWSEHQLOR---LQYRKLEMEQOAAFLV 560
DB 483 FTSFHLMDMIDGSAN---NSFVKVHLEKORLPGESEKRSYLUKAKYAKTEKDQIKM 538
OY 561 GSATPRYLTLAS-NHSMKMGHPGRIQMLSFAGEPLPONSSMARGFSMERQOLAVTQK 619
DB 539 SLXDPRYFHLVNPNRSLRGNPAGYKLVPGGNASLSLDHDDPPQMGCAFTNNGIWTTR-- 596
OY 620 EEEPPSSSVFNQNDPMA-----PTVD-FSDFINNETIAGDLVAVMTAGFLHIP 667
DB 597 -----YNRSEQWAGGLMTYQSRGEDTLQVWSD--RDRSIEKNDIYLVMTLGFHHPV 645
OY 668 HAEDIPNTVTVGNGVGFLLRPYFEDDP-----SFYSAD 702
DB 646 COEDFPMPTIAS--SEELKPVNFESNPVLGISPFEXD 683
RESULT 6
ABB92082
ID ABB92082 standard; Protein; 759 AA.
XX
XX ABB92082;
AC
XX 31-MAY-2002 (first entry)
DT
XX Herbicidally active polypeptide SEQ ID NO 1293.
DE
XX Herbicidal; plant; agriculture; herbicide.
XX
XX Arabidopsis thaliana.
OS
XX
XX WO200210210-A2.
XX
XX 07-FEB-2002.
XX
XX 28-AUG-2001; 2001WO-EP09892.
XX
XX 28-AUG-2001; 2001WO-EP09892.
XX
XX (FARB ) BAYER AG.
XX
XX Tietjen K, Weidler M;
XX
XX WPI: 2002-269010/31.
XX
XX Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plant with nucleic acid or amino acid sequences from non-plant
XX organisms -
XX
XX Claim 5; SEQ ID NO 1293; 261pp + Sequence Listing; English.
XX
XX The invention relates to identifying target proteins
XX (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX aligning and comparing nucleic acid or amino acid sequences from plant
XX with nucleic acid or amino acid sequences from non-plant organisms using
XX suitable search parameters, where plant sequences having an E-value
XX greater by a factor of 3 than the E-value of most similar non-plant
XX sequences are selected. The polypeptides or nucleic acids encoding them
XX are useful for identifying modulators. The identified modulators are
XX CC
XX
XX Sequence 759 AA:
SQ
Query Match 7.5%; Score 302.5; DB 23; Length 759;
Best Local Similarity 22.5%; Pred. No. 4.6e-21;

```


XX 05-SEP-1994; 94JP-0211526.
 XX 05-SEP-1994; 94JP-0211526.
 XX (Toku) TOKUYAMA SODA KK.
 XX WPI: 1996-203152/21.
 DR N-PSDB; AAT14174.
 XX
 PT Gene encoding amine oxidase - used in formation of hydrogen peroxide
 PT from butyl amine, or benzylamine by oxidative deamination
 PS Claim 3; Pages 9-11; 12pp; Japanese.
 CC The present sequence is the A. niger strain M-62 amine oxidase
 CC (AO), which is used in the formation of hydrogen peroxide from
 CC butyl amine, or benzylamine by oxidative deamination. The AO gene,
 CC the sequence of which was determined by sequencing a restriction
 CC enzyme map of an AO gene contg. cDNA, can be used for the
 CC recombinant prodn. of AO.
 XX
 SQ Sequence 670 AA:
 Query Match 5.3%; Score 214.5; DB 17; Length 670;
 Best Local Similarity 22.3%; Pred. No. 3,7e-12;
 Matches 168; Conservative 86; Mismatches 301; Indels 197; Gaps 37;
 34 HPOSOLFADLSRELVAVRFLTORLPGLVDAOARPSDNCVSEVLQPPKA----- 88
 4 HP-----LALSEETIANVILAQHPNTVIDPRE-----IYLSEPKAQLLEF 48
 89 AALAH---IDRGSPPARALA---IVFGROP-----QPNVSELVVGRLPHPS 131
 49 LALEHSGRLSPSPRRLALCOYDVLGNDRIPESESVVDVGRQVQHRVVGKEHHAS 108
 132 YMR---DYVER-HGRLPRHRRVLFQETL-DIDOMIFNRELPOA-----SGL--- 175
 109 LTLSEDTLVERCFASP-----LFQKALAFED-----LPEGEVVIIEPGGADYV 153
 176 -----LHCCFYKHKRGN-----LVTTMTAPRGLOGSDRATWFGLYNIGS 216
 154 EKKRRYFGICFATDKKKNPNDAFYRLPLFVMDGATRGDSGSPSR-----HCKG 208
 217 AGFELHNVGLLELVNHRKALDPARWTIQKVEYQGRYDSLQLEAFAGLVNVVLLPDNG 276
 209 EGLTEQFKRDIIGHCDS-----YVELL 234
 277 TCGSWSLKSPVPPGPAPPLQFYPOGPRFSV-QGSRVASSLWTSFGL---GAFSGPRIF 331
 235 PGGTRKDLKPLN-----VVOEGPSPFRTESLSLWEMOKWRFRVAENPRRCYHSOTSW 287
 332 DVAFPOGRLVYEISLOALAIYGGNSPAATTRVVD-GFGMGVYTTPLRGRVCPPLAT 390
 288 ---YDGSVILRLSVSEMTVPYADPRPPHKKQAFDGDGGGMANNLSTGCCCLAVIK 344
 391 YVDMHFLLESQAQKTINDATCFVEONOGRLPLRRHSDLYSHYEGFLAETV---LVYRSM 446
 345 YFPAVVMGADGSAKKMNAICLHDNDGIGMK-----HSWMRGRAVVFVHRELVAQFI 398
 447 STLNDYVMDYVFHPSGAIEIRYATGYSSAFLGATGKY---GNQVSE---HTLG 498
 399 ITLANEYIYAYKFDQSGGIGTRVACHGYLE-----RCQHRWCQOQRYVQRQMRVVG 451
 499 ---TVHNSAHFKVDLVAGLENNVMAEDVFPMAVPMWSEHQLQLOTRKLEHEEQ 555
 452 PEPPAHLCAHRP---GIYGNNSVQVEESHVPMAVNTNGVFYVNT-----ETWER 503
 556 AAFVLSATP---RYLYIASNH-----SNKGHPRGYRIQWLSPAGEPLPONSMAFGFS 607
 504 AGFF--DAPELNTVAMVNPHKKNPISQK---PVGYKFIPLAQORLLADNSIQARRAQ 558
 608 WERYQLAVTORKEEPESSSVFN-QNDPMAFTVDSPT-NNETIAGKDLVAVMTAGFLH 665

DB 559 FAQHWWVTKYRQGEIYAGRYTLQSGEEIEGV--SDAVRGDSVDTDVVWSTFGITH 616
 QY 666 IPHAEDIPNTVTGNGVGFFLRPVNFDEPDS 697
 DB 617 NPEVEDMP--VMEVEIFQLMIRPADFTTANPS 646
 RESULT 9
 AAB42952
 ID AAB42952 standard; Protein; 863 AA.
 AC AAB42952;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF2716 polypeptide sequence SEQ ID NO:5432.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 KW
 XX Homo sapiens.
 XX
 PN NO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000MO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Leach M;
 PT WPI: 2000-602362/57.
 DR N-PSDB; AAC77161.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 11; Page 4615-4617; 5507pp; English.
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,

KM bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;
 KW thiazolidine; bleomycin; microbial metabolite; sugar.
 OS Streptomyces verticillatus.
 XX WO200040704-A1.
 PN 13-JUL-2000.
 PD 06-JAN-2000; 2000MO-US00445.
 PE 06-JAN-1999; 990US-0115435.
 PR 05-FEB-1999; 990US-0118848.
 PR 05-JAN-2000; 2000US-0477962.
 XX (REGC) UNIV CALIFORNIA.
 PA
 XX Shen B, Du L, Sanchez C, Chen M, Edwards DJ.
 PI WPI: 2000-465974/40.
 DR N-PSDB; AAA58471.
 XX
 PT New bleomycin gene cluster components useful for peptide and/or
 PT polyketide metabolites, especially bleomycin, production and for
 PT chemically modifying biological molecules -
 XX
 PS Disclosure: Page 97-137; 162pp; English.
 CC AAB07556-78 represent proteins encoded by open reading frames (ORFs)
 CC 8 to 30 of the BLM (Bleomycin) gene cluster. The proteins encoded
 CC by the gene cluster are useful for producing peptides and/or polyketide
 CC metabolites, especially bleomycin or bleomycin analogues. They are
 CC also useful for chemically modifying biological molecules to produce
 CC branched methyl groups, and for coupling amino acids and fatty
 CC acids. They may be reacted with an apo-carrier protein and coenzyme A
 CC to produce a holo-carrier protein. The BLM gene cluster or catalytic
 CC domains can be used individually or collectively to produce
 CC thiazolidine, thiazoline, bithiazoline and bithiazoline-containing
 CC microbial metabolites. The BLM gene cluster may also be used to produce
 CC sugars.
 CC
 SQ Sequence 2675 AA;
 Query Match 2.9%; Score 115; DB 21; Length 2675;
 Best Local Similarity 18.9%; Pred. No. 0.52;
 Matches 172; Conservative 107; Mismatches 321; Indels 310; Gaps 42;

DB 1628 -----AERLREYGVDRDFESLGGDSVRALGVTAALRAAGLPVTVDILRLPTVA 1678
 QY 387 YLATYVDMH-----FLI--ESQAPKTRDAFCVEQNOGLPLRRHNSDLXS 430
 DB 1679 ALARHADERADRRPARQETPPGFALCPENAGPGLGDAVPMSMAQRAVLFHRDHPGYE 1738
 QY 431 HYFGGLA-----ETVLVRSMSSTLLN-YDYV---WDTVFHPSGA-----IEIRY 471
 DB 1739 VYVTSVAVSTPLDRTRLAAAVRLDRHAVALRSSFDLVSHPEPTQLWTHLPTELEV--- 1795
 QY 472 ATGYISSAFLFG-----ATGKYGNVSEHTIGTVTH---SAHFVVDVAGLEWYMA 522
 DB 1796 ---VESDPAGFDANLHAEKKRPDVGTCPLARTAHDAAGAGFRITVSSFDLQGCYVA 1851
 QY 523 EDWVEV-----PMAVP-----WSPE-----HOL 540
 DB 1852 TVLTLLRDYWSALRGAPLSLPAPASREFVALERAQDHAPHREFWRELGAARPHLP 1911
 QY 541 QRLQVT-----RKLEMEQQA-----FLYG-----SATP 565
 DB 1912 PRRPVPPPGPDGIRQHRHYVPEDTVAKGISALAGELGVGLKHYLLGVHLRVVALSGDP 1971
 QY 566 RYLLXASNHSNKKGH-----PRGYRIQMLSPAGEPLPONSSMARGFSMERYO 612
 DB 1972 DVTAVETHGRLEHRHDGDRVLGVFNILPLRQRYDGGSNADLAAHAAARAGTEYRRTP 2031
 QY 613 LAVTORKEEPESSSVFNQNDPAPTPVDSDFINNETIAGKDLVAVWTAQFLH1PHAEDI 672
 DB 2032 LAQAQR---DHGAAGLED-----TLFVFHFLYLRALADLDGMA---VEDLAPAPQTVY 2079
 QY 673 PNT-----VTVNGVGFILR--PYNFDEDBEFTYSADSIYFFGGDODACAEVNPALCP 724
 DB 2080 PLTAHNRVATDGGGRLLESDPREPPDQVAEFNA---YXRR-----AL 2122
 QY 725 QAAACAPDLR 734
 DB 2123 RAADADPHRP 2132
 RESULT 12
 AAR32999
 ID AAR32999 standard; protein; 435 AA.
 AC AAR32999;
 XX
 DT 14-JUN-1993 (first entry)
 XX
 DE Rat choline kinase.
 XX
 KW CK; epitope: tumour; diagnostic; antibodies; distribution; liver;
 KW cranial nerve disease.
 XX
 OS Rattus rattus.
 XX
 FH Key Location/Qualifiers
 FT Peptide 139..225
 FT Peptide 278..327
 FT Peptide 328..360
 FT Peptide 398..412
 PN JP05015367-A.
 XX
 PD 26-JAN-1993.
 XX
 PF 17-SEP-1991; 91JP-0262516.
 XX
 PR 14-SEP-1990; 90JP-0242528.
 XX
 PA (MITK) MITSUI TOATSU CHEM INC.
 XX
 DR WPI: 1993-070176/09.
 DR N-PSDB; AAQ37687.
 XX

PT Rat choline kinase having specified DNA sequence - useful for
PT diagnosing cranial nerve diseases, liver diseases, tumours, etc.
XX
XX
PS Claim 1, Page 21, 26pp: Japanese.

XX The choline kinase (CK) gene may be obtd. from rat brain, human
CC brain, human liver, chicken liver, egg, soy bean, etc. The rat CK
CC gene is shown, and can be isolated by PCR using oligonucleotides
CC corresp. to partial CK sequences as primers. The oligonucleotides
CC or the amplified DNA fragments, can then be used as probes to detect
CC CK coding sequences. Rat CK may be purified by MMA-agarose affinity
CC chromatography. A large amt. of CK may be produced and antibodies
CC against CK may be obtd. for use in detecting the distribution of CK
CC in tissues or cells, and the diagnosis of cranial nerve disease.
CC liver disease and tumours.
XX

SO Sequence 435 AA;

Query Match 2.7%; Score 109; DB 14; Length 435;
Best Local Similarity 21.5%; Pred. No. 0.11;

Matches 79; Conservative 41; Mismatches 150; Indels 98; Gaps 14;

OY 249 GRYDLSLAQLAOFKAGLVNVLIPDNGTGGWSLKSVPPEPPAPLPLOFPQGPFFSVQG 308
DB 33 GQORDAAGELSKDGLGSRGSLALPPPPP-----PLPLPPPPSPPLADEQPPAPR----- 82
OY 309 SRVASSLTFFSGFGAFSGPR-----IFDVRFQGERLYVEISLOEALAYGNSPAMMT 363
DB 83 TRRAYLMCKEFLFGAMGRLBEDOFHIVINGLNNMLFQCSLPSIASV-GDEPRKYL 141
OY 364 RYVDGFGMGKTYTPPLRGVDCPYLATYVDWHFLESQAPKTIRDAFCVEQ---NQGLP 420
DB 142 RL-----YGAIIKMGAEAMVLESVMFALIAERSLGP---LYGIFFQGRLEQRIIP 188
OY 421 LRRHSD--LXSHYFGAGAEVLYVRSMTLLNDYVDYVDFHSGAIEIRPYATGYISS 478
DB 189 SRRLDFEELCLPDISAELAEKMAFHHGKMPFNKEPKV----- 226
OY 479 AFLFGATGKYGNOVSEHTLGTVHTSAHFKYDLVAGLENNVMADNVFVPAVWSPFH 538
DB 227 --LFGTMEKYLNOYL-----LKFSEARVQQLHK-----FLSYNPLELEN 266
OY 539 QLOQLQVTRK-----LLE---MEQAAFLVGSATPRYLTA---SNHSH 576
DB 267 LRSLLQYTRSVFVCHNDCEGNILLLEGQENSEKOKLMLIDFEYSYVNGFDIGNHFC 326
OY 577 KMGHPRGY 584
DB 327 EMTDYTYT 334

RESULT 13
ABG3153
ID ABG3153 standard; Protein: 446 AA.

XX ABG3153;

XX 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #13144.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.

OS WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.
XX (HSE-) HXSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.
DR N-PSDB: AAS77340.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX

PS Claim 20; SEQ ID No 43512; 103pp: English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX

SO Sequence 446 AA;

Query Match 2.7%; Score 106.5; DB 22; Length 446;
Best Local Similarity 21.3%; Pred. No. 0.2;

Matches 65; Conservative 39; Mismatches 98; Indels 103; Gaps 16;

OY 333 VRFGGERLYVEI-SLOELATAYGNSPAMTRY-----YGGGGMGKTYTPPLRGV 383
DB 211 LKAEGERAVYQIMOLQKLVNMGSD---LTRRYGQKVNKLTLLHGFGSPCPNRDGTIGRG- 265
OY 384 DCPY--LATYVDHMFLESQAPKTIRDAFCVEQNGOGLRLRHSGL-----YSHFGG 435
DB 266 GCTFCNNAFSD-----EAQOHRSAIEOLA-----HOANLVNRAKRYLGTFQA 308
OY 436 LAET---VLVVRSMSTLLNDYVDYVDFHFGAIEIR-FYATGYISSAF-----LFGAT 485
DB 309 YTSFAEYQVLRSM-----YQAAVEHTGTGKVGITGFCGCVSNAAYAVPELCAV 360
OY 486 GKYGNOVSEHTLGTVHTSAHFKYDLVAGLENNVMADNVFVMAVWSPFHQLOQLQV 545
DB 361 PFYGRQAPTDVAKTEAPLLHFAELDTFINEG-----PAYE-AALKA 403
OY 546 TRKLEMEEQAAFLVGSATPRYLTAASNSHMKWGHPRGYRIQMLSPAGEPLPOJSSNARG 605
DB 404 NNKYYE-----AYITPGVNH-----GFHNDSTPRIDKSAAD 434
OY 606 FSWER 610
DB 435 LAMOR 439

RESULT 14
AAE23407
ID AAE23407 standard; Protein: 617 AA.

AC AAE23407;
 XX
 DT 27-AUG-2002 (first entry)
 XX
 DE Lysine oxidase proenzyme mut3.
 XX
 XX Amino acid oxidase; insect infestation; gene therapy; boll weevil; BWV;
 XX corn rootworm; CRM; insecticide; wireworm; WM; Colorado potato beetle;
 KW CPB; lysine oxidase; tetanactam synthase; enzyme; E.C. 1.4.3.2;
 XX
 XX Unidentified.
 OS
 XX
 XX Key Location/Qualifiers
 FH Region 68..86
 FT /note= "Protease sensitive region"
 FT
 FT Misc-difference 70
 FT /note= "Wild-type Ala substituted with Gly"
 FT
 FT Misc-difference 71
 FT /note= "Wild-type Leu substituted with Gly"
 FT
 FT Misc-difference 72
 FT /note= "Wild-type Leu substituted with Gly"
 FT
 FT Misc-difference 73
 FT /note= "Wild-type Lys substituted with Gly"
 FT
 FT Misc-difference 74
 FT /label= Unknown
 FT /note= "Wild-type Glu substituted with Xaa"
 FT
 FT Misc-difference 75
 FT /label= Unknown
 FT /note= "Wild-type Ala substituted with Xaa"
 FT
 FT Misc-difference 76
 FT /label= Unknown
 FT /note= "Wild-type Pro substituted with Xaa"
 FT
 FT Misc-difference 77
 FT /label= Unknown
 FT /note= "Wild-type Arg substituted with Xaa"
 FT
 FT Misc-difference 78
 FT /label= Unknown
 FT /note= "Wild-type Ala substituted with Xaa"
 FT
 FT Misc-difference 79
 FT /label= Unknown
 FT /note= "Wild-type Glu substituted with Xaa"
 FT
 FT Misc-difference 80
 FT /note= "Wild-type Glu substituted with Gly"
 FT
 FT Misc-difference 81
 FT /note= "Wild-type Glu substituted with Gly"
 FT
 FT Misc-difference 82
 FT /note= "Wild-type Leu substituted with Gly"
 FT
 FT
 XX US6372211-B1.
 PN
 XX
 XX 16-APR-2002.
 PD
 XX
 XX 21-APR-1998; 98US-0063733.
 PF
 XX 21-APR-1997; 97US-044504P.
 PR
 XX (MONS) MONSANTO TECHNOLOGY LLC.
 PA
 XX
 XX Isaac BG, Greenplate JT, Purcell JP, Romano CP;
 PI
 XX
 XX WPI; 2002-424731/45.
 DR
 XX
 XX Composition for controlling coleopteran insect infestation of plants,
 PT such as Diabrotica, Melanotus, Leptinotarsa, or Anthonomus infestation,
 PT comprises a lysine oxidase enzyme and a tetanactam synthase enzyme -
 XX
 XX Example 13; Column 89-94; 77pp; English.
 PS
 XX
 XX The invention relates to composition and methods for controlling
 CC coleopteran insect infestation of plants, by co-expressing an amino
 CC acid oxidase and a second enzyme that provides insecticidal activity
 CC when present in a mixture with the amino acid oxidase. Nucleic acid

CC sequences encoding these enzymes are used in gene therapy. The
 CC composition is used to control insect infestation of plants. It is
 CC used to control coleopteran species Diabrotica, Anthonomus, Melanotus,
 CC or Leptinotarsa. It is especially used to control boll weevil (BWV),
 CC corn rootworm (CRM), wireworm (WM) or Colorado potato beetle (CPB).
 CC The composition can also be used to transform plants that can express
 CC the enzymes of the composition. The present sequence is a lysine
 CC oxidase (E.C. 1.4.3.2) proenzyme mutant used in the exemplification
 CC of the invention.
 CC
 XX
 XX Sequence 617 AA;
 SQ
 Query Match 2.6%; Score 106; DB 23; Length 617;
 Best Local Similarity 22.8%; Pred. No. 0.38;
 Matches 79; Conservative 40; Mismatches 141; Indels 86; Gaps 18;
 OY 44 LSREELTAVMRFTQRLG--PGLVDAQAAR-----PSDNCVFSVELQPPKAA 90
 DB 18 LIREKXAKELNITERLGEVPIPPNEGFLGGYSHDLPSPDPLYSISK-----PGGG 73
 OY 91 LAHLDGSPPPAREALAIYFFGRQPPQPNVSELYG-----PLPHSYMDVVERHGR 144
 DB 74 XXXXXXGCGPPRK--VCIVGAG-----VSGLYIAMILDLPKIPNLYDIFESSRTRGR 125
 OY 145 LPYHRRPVLQGEYLDIDOMFNRRLPOAGSLHHCFCYKRGMLVMTTAPRGLSGDR 204
 DB 126 LYTHHTDAKHDYDIDGAMY-----PDIPSKRTFNLFKRTGAPLI-----KYLDGEN 175
 OY 205 ATMFGLYNNISGAGFELHVGLELVN-----HKALDPARMTIQKVFYGRYYDSLA 256
 DB 176 TP--QLYNN-----HFAKGVSDPYMVSANGTVDPDDVDVSVEKLCQAAR--GYREKLA 227
 OY 257 QLEAGPEAGLVNVLITDNTGGSWSLKSVPYPPGAPPLDYPQGFPSVQGSRYVSSSLW 316
 DB 228 E---DFDKGFDELVDVDMIT--REYLKRGPGGEAKRYDF-----FAIQMMETON--- 273
 OY 317 TFSRGLGAFSGPRIFDVRFQGERLVEISLQ-----PALAIYGNS 357
 DB 274 -----TGTLFDQAF--SESVYDSFDDNPFPKPMYCLEGSTS 309
 RESULT 15
 ID ABB67526 standard; Protein; 825 AA.
 AC ABB67526;
 XX
 XX 26-MAR-2002 (first entry)
 DT
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 29370.
 XX
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 XX Drosophila melanogaster.
 OS
 XX
 XX WO200171042-A2.
 PN
 XX
 XX 27-SEP-2001.
 PD
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 PF
 XX
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 XX (PEKE) PE CORP NY.
 PA
 XX
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX
 XX WPI; 2001-656860/75.
 DR
 XX N-PSDB; ABL11629.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 12:02:34 ; Search time 20 Seconds

(Without alignments)
1094.533 Million cell updates/sec

Title: US-10-081-408-2_COPY_20_763

Perfect score: 4010
Sequence: 1 LVCVLLVGRGSGDGPSPQLP.....OMACAPDLPALFSGHGFSSHN 744

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 2942292 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents AA.*
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/6C.COMB.pep.*
7: /cgn2_6/ptodata/2/1aa/6D.COMB.pep.*
8: /cgn2_6/ptodata/2/1aa/6E.COMB.pep.*
9: /cgn2_6/ptodata/2/1aa/6F.COMB.pep.*
10: /cgn2_6/ptodata/2/1aa/6G.COMB.pep.*
11: /cgn2_6/ptodata/2/1aa/6H.COMB.pep.*
12: /cgn2_6/ptodata/2/1aa/6I.COMB.pep.*
13: /cgn2_6/ptodata/2/1aa/6J.COMB.pep.*
14: /cgn2_6/ptodata/2/1aa/6K.COMB.pep.*
15: /cgn2_6/ptodata/2/1aa/6L.COMB.pep.*
16: /cgn2_6/ptodata/2/1aa/6M.COMB.pep.*
17: /cgn2_6/ptodata/2/1aa/6N.COMB.pep.*
18: /cgn2_6/ptodata/2/1aa/6O.COMB.pep.*
19: /cgn2_6/ptodata/2/1aa/6P.COMB.pep.*
20: /cgn2_6/ptodata/2/1aa/6Q.COMB.pep.*
21: /cgn2_6/ptodata/2/1aa/6R.COMB.pep.*
22: /cgn2_6/ptodata/2/1aa/6S.COMB.pep.*
23: /cgn2_6/ptodata/2/1aa/6T.COMB.pep.*
24: /cgn2_6/ptodata/2/1aa/6U.COMB.pep.*
25: /cgn2_6/ptodata/2/1aa/6V.COMB.pep.*
26: /cgn2_6/ptodata/2/1aa/6W.COMB.pep.*
27: /cgn2_6/ptodata/2/1aa/6X.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	106	2.6	617	4	US-09-063-733A-58
2	102.5	2.6	541	4	US-09-158-767-19
3	102.5	2.6	541	4	US-09-158-767-20
4	101	2.5	617	4	US-09-063-733A-46
5	100.5	2.5	1253	4	US-08-506-296B-14
6	100	2.5	617	4	US-09-314-242-2
7	95	2.4	617	4	US-09-063-733A-57
8	94.5	2.4	1751	4	US-09-135-574A-44
9	94.5	2.4	15281	2	US-08-471-119A-2
10	93.5	2.3	485	5	PCT-US94-03437-8
11	93.5	2.3	486	1	US-07-737-071A-2
12	93.5	2.3	486	1	US-08-022-096-2
13	93.5	2.3	486	3	US-08-445-463B-8
14	93.5	2.3	486	3	US-08-445-463B-8
15	93.5	2.3	486	3	US-08-445-463B-8
16	93	2.3	2972	4	US-08-044-857D-8
17	93	2.3	3118	4	US-09-579-181-1
18	92.5	2.3	448	2	US-08-078-311-12
19	92.5	2.3	448	2	US-08-460-402-12
20	92.5	2.3	493	2	US-08-078-311-7
21	92.5	2.3	493	2	US-08-460-402-7
22	92.5	2.3	635	1	US-08-184-327A-4
23	92.5	2.3	635	2	US-08-078-311-1
24	92.5	2.3	635	2	US-08-460-402-1
25	92.5	2.3	635	5	PCT-US95-00670-4
26	90.5	2.3	1436	2	US-08-652-971-2
27	90.5	2.3	1436	2	US-08-991-258A-2

28	90.5	2.3	1436	2	US-08-769-399-2	Sequence 2, Appl
29	90.5	2.3	1436	3	US-08-991-953A-2	Sequence 2, Appl
30	90.5	2.3	1467	4	US-09-523-656-38	Sequence 38, Appl
31	90	2.2	617	4	US-09-063-733A-56	Sequence 56, Appl
32	89.5	2.2	623	1	US-08-332-838-2	Sequence 2, Appl
33	89.5	2.2	630	4	US-09-228-986-71	Sequence 71, Appl
34	89.5	2.2	911	4	US-08-460-269C-4	Sequence 4, Appl
35	89	2.2	688	4	US-08-973-273-26	Sequence 26, Appl
36	89	2.2	747	4	US-08-973-273-3	Sequence 3, Appl
37	88.5	2.2	379	3	US-09-028-934-36	Sequence 36, Appl
38	88.5	2.2	512	1	US-08-462-484-6	Sequence 6, Appl
39	88.5	2.2	512	1	US-08-441-147-6	Sequence 6, Appl
40	88.5	2.2	512	5	PCT-US95-07536-6	Sequence 6, Appl
41	88.5	2.2	583	4	US-09-401-476-2	Sequence 2, Appl
42	87.5	2.2	353	4	US-08-838-151A-49	Sequence 49, Appl
43	87.5	2.2	485	3	US-08-445-463B-2	Sequence 2, Appl
44	87.5	2.2	485	3	US-08-445-463B-2	Sequence 2, Appl
45	87.5	2.2	485	4	US-08-044-857D-2	Sequence 2, Appl

ALIGNMENTS

```
RESULT 1
US-09-063-733A-58
; Sequence 58, Application US/09063733A
; Patent No. 6372211
; GENERAL INFORMATION:
; APPLICANT: Isaac, Barbara G.
; APPLICANT: Greenplate, John T.
; APPLICANT: Purcell, John P.
; APPLICANT: Romano, Charles P.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROLLING
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Arnold White & Durkee
; STREET: PO Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/063,733A
; FILING DATE: 21-APR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Patterson, Melinda L.
; REGISTRATION NUMBER: 33,062
; REFERENCE/DOCKET NUMBER: MOBT:022
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1400
; TELEFAX: 713-787-1440
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 617 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; US-09-063-733A-58
;
; Query Match 2.6%; Score 106; DB 4; Length 617;
; Best Local Similarity 22.8%; Pred. No. 0.071;
; Matches 79; Conservative 40; Mismatches 141; Indels 86; Gaps 18;
;
; 44 LSEELTAMRFLTOLG--PGVDAAR-----PSDNCVFSELDLPKAA 90
; 18 LIREKAKELNITLRLGEVPIPPNEGFLGGYSHDNLPSDPLYSTIK---PGGGG 73
```

QY	91	LAHLDRGSPPARALALIVFGRRGPOGNVSLVYG-----PLPHSEYMDVYERHGGP	144
Db	74	XXXXXXXXGGGPPRK--VCIVAG-----VSLGTAMLLDKITPNLTYYDIFESSRFGGR	1255
QY	145	LPHYRRPYLFGQYLDIDOMIFNRELPOASGLLHHCCFYKRRGNLYMTTAPRGLOSGDR	204
Db	126	LYTHHFDTAKHDYDIDGAMRY---PIPSMKRFFNLFKTRGMPLI-----KYLDDGEN	175
QY	205	ATWEGLYYNIGSAQGFLLHNVLELLVY-----HKAALDPARMTIDKVYFGRYVDSLA	256
Db	176	TP--QLYNN---HFFAKGVSDPIWVSANGVTPDDVYDVSVEGKLDQAF--GYKKKLA	2277
QY	257	QLEAQFEAGLVNVYLLIPDNGTGGSMSSLSKPYPPGAPAEPLFYPOGPRFYSVQGSKVASLSW	316
Db	228	E---DFDKGFDEMLVDDMT--REYLKRGCPKEAEKYPD-----FALQMMETQN---	273
QY	317	TFSTGLGAFSGPRIFDVRFGQERLYVETSLQ-----EALAIYGSNS	357
Db	274	-----TGTNLEDAAF--SESVIIDEFDPNTPKPEKTCLEGGTS	309

```

RESULT 2
US-09-158-767-19
: Sequence 19, Application US/09158767A
: Patent No. 6180363
: GENERAL INFORMATION:
: APPLICANT: Batard, Yannick
: APPLICANT: Durst, Francis
: APPLICANT: Schaik, Michel
: APPLICANT: Werck-Reichardt, Daniele
: TITLE OF INVENTION: RECODING OF DNA SEQUENCES PERMITTING
: TITLE OF INVENTION: EXPRESSION IN YEAST AND OBTAINED TRANSFORMED YEAST
: FILE REFERENCE: A33000
: CURRENT APPLICATION NUMBER: US/09/158,767A
: CURRENT FILING DATE: 1998-09-23
: EARLIER APPLICATION NUMBER: FR 97-12094
: EARLIER FILING DATE: 1997-09-24
: NUMBER OF SEQ ID NOS: 20
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 19
: LENGTH: 541
: TYPE: PRM
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Altered sequences
US-09-158-767-19

```

Query Match	2.6%	Score 102.5;	DB 4;	Length 541;
Best Local Similarity	19.6%	Pred. No. 0.13;		
Matches 109;	Conservative 60;	Mismatches 205;	Indels 183;	Gaps 22;

```

0Y 151 PVLPEYEDIDOMIFENRLEPQASLLHNCCP-----YKNRBNLYVTMTTARPGLOSDBRT 206
Db 40 PELVONHADMEHWMLG-NLRRAGSTYOTCTIFAPGVARRG-GLVYTYCDPRNLEHWLKAR 97
0Y 207 -----WFLVYVYNSGAGF-----LHNVLELLVNHKALDARMTIOKVY 247
Db 98 FDNPEKGFPMHGVFREDLGDIDFNSDDTMYAOKRTALEFTTTRTLTAMSRVYSRI-- 155
0Y 248 OGRYDLSLAOL---EAOGEAGLVVWVLIPIBDTGSGMSLSPVPRGPAPLOFYPOOPRF 304
Db 156 HQRLLPILADAKKGAOVDDODLLRLTFNIGCLG-----KDEPLAQGLPENEFRA 209
0Y 305 SVQSGRVASSLWTSF-----GLGAFSGPRIFDVRFQGERLYEISLQEALEY 353
Db 210 SAFDPAETATLNRFIFPEFLRCKRWLGLG-----METTLTSSMAHV 251
0Y 354 GNSPAAATTFYVQGGFQMGKYTTPLR-----GVDC 385
Db 252 DOYLAAYLTKKKRLELAGNGKCDTPAHTHDDLSPFMKKGTSIDESLOHVALNPLACRDT 311
0Y 386 PYLATYVDMHFLLESQAD-----KTIKRAFVCEQNOGL-----PLRRHNSDLYSHY 432

```

```
Db      312 SSV- -LSMFVLWSTHPAVERKIVRELCSYLASRCAGHDPALMAIEPFEEELDRAYVL    369
Qy      433 FGGIAENVLYVRNS-----SRLTANDDYMDVDVPHPSG-AIEIRFANGYSASFLLPATGK    487
Db      370 KAALSEIKRLPSPPEBSKHVVADILPDGIFFPAGSSVTYSISGRMG-----          420
Qy      488 YGNOVSEHTLCTVHTSHAHFKVDLDVAGLENWAEDMFEVPMAVWPSPHEOLORLOYTR    547
Db      421 -----TWGED-----CLEERPE---RWLSADG    439
Qy      548 KLEMEBOAEFLVGSAATPRYLULASNHSNKWGHRGYRIOMLSFAGEPLPONSSMARGFS    607
Db      440 TKFEQHDSYKKVAFNAGRVC-----GRDLAV-LQMKNTAGSVLLRH-----          481
Qy      608 WERYOLANTORKKEEPS   624
Db      482 --RLTVAPGHRYEOKMS 496
```

```

RESULT 3
US-09-158-767-20
; Sequence 20, Application US/09158767A
; Patent No. 6180363
; GENERAL INFORMATION:
; APPLICANT: Batard, Yannick
; APPLICANT: Durst, Francis
; APPLICANT: Schalk, Michel
; APPLICANT: Werck-Reichardt, Daniele
; TITLE OF INVENTION: RECORDING OF DNA SEQUENCES PERMITTING
; TITLE OF INVENTION: EXPRESSION IN YEAST AND OBTAINED TRANSFORMED YEAST
; FILE REFERENCE: A32000
; CURRENT APPLICATION NUMBER: US/09/158,767A
; CURRENT FILING DATE: 1998-09-23
; EARLIER APPLICATION NUMBER: FR 97-12094
; EARLIER FILING DATE: 1997-09-24
; NUMBER OF SEQ. ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Altered sequences
US-09-158-767-20

```

Query Match	2.6%	Score 102.5;	DB 4;	Length 541;
Best Local Similarity	19.6%;	Pred. No. 0.13;		
Matches 109;	Conservative 60;	Mismatches 205;	Indels 183;	Gaps 22;

```

QY 151 PVLFEYDIIDIOIMFENRLEPQASLLHNHCCF---YKHGRNLVTMTTARGLQSGGRAT 206
Db 40 POLYONADHMENIAG-NLIRARAGTYOTCIFAVPGVARRG-GLVTYTCSDPRNLEHYLAKR 97
QY 207 -----WGLYYNISGAGFF-----LHNYGELLYVNHKALDPAKRTIQCVFY 247
Db 98 FDNVKGPFMGHGVPRDLLDGDIDFNSDGDWTIAQKRTAALETFTTFTLTAMSRVWSRI-- 155
QY 248 QGRYVDSIALOC---EAGFEAGLVNVVLLIPNDGTGSGWSLKSVPVPRGAPRLOEYPOGPR 304
Db 156 HGRLLPILADAKAKAGQVUDLDLRLTFLDNIGSLAG---KDEPLAQLGPRENEFA 209
QY 305 SVQSGRVASSLMTBSF-----GLAFSGPRIFDVAFQGERLUYELISQENALIT 353
Db 210 SAEFDRAETATLNRLEFPEELRCKKMWGLG-----METTLSSMAHY 251
QY 354 GGNSPAAMTTRVYDGGFEGKUTYPLTR-----GVDC 385
Db 252 DQYLAAYIKKKRLELAAGNGKCDPAATHDDLRSFMKGSYSDSLQHYALNFIAGRDT 311
QY 386 PLATLYVWHLELBSQAP---KIIRDAFCYFEQNOGL-----PLRHHNDLXSHY 432
Db 312 SSVA--LSMFVWLSTRAVBRKILVYRELCSVLASRGANDRALMLABPFFFEELDRVLVL 369

```

```

QY 488 YGNVSEHTJGTVHTSHAHKFKVDLDVAGLENNVAEDMEFVPAVAWPSPEHQLQLOQVTR 547
Db 421 -----WGGE-----CLERPRE-----RWLSADG 439
QY 548 KLEMEQALVGSATPRVLYILASHNSKMGHRGRYRIOMLSRAGEPLQONSSMAGFS 607
Db 440 TKFQHOHSYKRVAAFNAGPRVCL-----GKDLAY-LQMKNIAGSVLLRH----- 481
QY 608 WERYOLAVTORKEEPS 624
Db 482 --RLTVAPGHRVDEQMS 496

```

```

RESULT 4
US-09-063-733A-46
; Sequence 46: Application US/09063733A
; Patent No. 6372211
;
; GENERAL INFORMATION:
; APPLICANT: Isaac, Barbara G.
; APPLICANT: Greenplate, John T.
; APPLICANT: Purcell, John P.
; APPLICANT: Romano, Charles P.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROLLING
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: PO Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/063,733A
; FILING DATE: 21-APR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Patterson, Melinda L.
; REGISTRATION NUMBER: 33,062
; REFERENCE/DOCKET NUMBER: MOBT.022
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1400
; TELEFAX: 713-787-1440
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 617 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
;
US-09-063-733A-46
Query Match 2.5%; Score 101; DB 4; Length 617;
Best Local Similarity 23.3%; Pred. No. 0.24;
Matches 81; Conservative 39; Mismatches 138; Indels 90; Gaps

QY 44 LSRELLTVAMRFLEIQRGLG-DGLYDAQAAR-----PSCNCFVSVELQLPPRAAA 90
DB 18 LIREYVAEELNLTIRLEGEVGIPIPPNCGREFLGCGYSHDNLPSPLYSSI-----KPAL 71
QY 91 LAHLDGRSP--PRAEALAIYFFGNGRPPNNSSELYVG-----PLPHRSYKRDVTVRRHG 142
DB 72 LKEADRAEELLPKRR--VCIVGAG-----VSGLYIAMDLDLKLIPNLTVDIFFESSRTG 123

```

```

QY 203 DRATFGLLYNISCAGFGLHHVGLLELVN-----HKALDPARMTIOKVYFOGRYDS 254
      :      :      :      :      :      :      :      :      :      :
Db 174 ENTP--OLLYN-----HEFAKGVSPRYVAVSANGTVEDDVDVDSYGEKIOAAF--GYKKEK 225
      :      :      :      :      :      :      :      :      :      :
QY 255 LAOLEAFEGAGLVNVLIPNNGTSGMSLKSVPYGPAPPLQFQPOSPPRSVSGSRVASS 314
      ||      :      :      :      :      :      :      :      :      :
Db 226 LAE--DPDGFEDELMVDVMTT--REYLKRGKFGKAPRYDF-----FAIOMETON- 273
      :      :      :      :      :      :      :      :      :      :
QY 315 LMTFSGIAGFSSGRIFDVRFQGERLYEYISLQ-----EALAIYGGNS 357
      :      :      :      :      :      :      :      :      :      :
Db 274 -----TGNTLFDQAF--SSSVYIDSPFDNPPTKPEWYCIEGGS 309

```

```

RESULT 5
US-08-506-296B-14
; Sequence 14, Application US/08506296B
; Patent No. 631265
GENERAL INFORMATION:
APPLICANT: Phillips, Greg
APPLICANT: Cunningham, Bruce A.
APPLICANT: Crossin, Kathryn L.
TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES
TITLE OF INVENTION: CONTAINING FIBRONECTIN TYPE III REPEATS AND METHODS OF USE
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute
STREET: 10550 NO. 6131265th Torrey Pines Road, TPC-8
CITY: La Jolla
STATE: California
COUNTRY: U.S.
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/506/296B
FILING DATE: 24-JUL-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: TSRI 488.0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 554-2937
TELEFAX: (619) 554-6312
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1253 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-506-296B-14

Query Match      2.5%; Score 100.5; DB 4; Length 1253;
Best Local Similarity 18.5%; Pred. No. 0.91;
Matches 160; Conservative 89; Mismatches 269; Indels 349; Gaps 40;

OY    4 VLVVRGGD-----GEPSLPHCPVSFSACP-WTHPQOSQLFA----- 42
       :||| |I| | | | | | | | | | | | | | | | | | | | | |
DB    299 ILKVEEDDDGECLAEISLSGARHAYVTVYEAAPYLHKPKSHLYGGETARLDCCVOYG 358
       :||| |I| | | | | | | | | | | | | | | | | | | | | |

OY    43 -----DLSEKETAAVRFRTLRGLGVDAQAARSDNCVFSE----- 81
       : ||| | | | | | | | | | | | | | | | | | | | | |
DB    359 RQPEVTWRINKIPTELAKDKQRYIQR---CALLSNVPQSPDTFTVTCLEARNRHGLLA 415
       : ||| | | | | | | | | | | | | | | | | | | | | |

OY    82 -----LQLPPKAAALAHIDRGSPPARDALAIYFFGKQPQPNSELVAGLPHPSTWMD 135
       : ||| | | | | | | | | | | | | | | | | | | | | |

```

```

Db 416 NAVIYVQIPAKILTDADNOTYMAVOGSTAYLLCKAFG-APVPSV-----458
QY 136 VTERHGGRPLPHRRPVLFOEYLDIDOMIF---NRELPOASGLLHCCFYKRGKRNLYTM 192
Db 459 -----QWLEDEDTTVLQDERFEFFYANG-----480
QY 193 TTAPRGLOSGRATWGLYUYNISGAGFELHHVGLLELVNHNALDPAWTKQVYQGRXY 252
Db 481 TLGIRLDQANDTGYFCLLANDON-----NTIMANKVKQATQIT-----521
QY 253 DSLAOLEAGFAGLVNVLIPDNGTGSWSLSKSPGPPAPLOFYPOGPRFSVQ--GSR 310
Db 522 -----QGRSTIEKKGR 534
QY 311 VASSLMTFSGLGAFSGPRIFD-----VROGE-RLVYEISLOBALAIYGNSPAAMTT 363
Db 535 V-----TFQOASDPSPLOPSTWGRGDLQSDKXYFIEDGR- VJHSL 581
QY 364 RYVDGFGMGKYTPPLTRGVDCPYLATYVDHFLLESQA-----PKTRDAFCV 412
Db 582 DYSD-----QGVYSCVASTELD-----VYESRAQLLVGSPGPRVRLVLSDLHL 625
QY 413 FEQVQ---GLPLRHNSDLXSHFGLAETVLYVRSKSTILNDYVDVYFHPSGALEI 468
Db 626 LTOQOVHVSMPADHNAPIEK-----YDIEFEDKEMAPE 660
QY 469 RFYATG-----YISSAFLEGATGKYG---NOVSEHMLGTVHT-HSAHF 508
Db 661 KMYSLGKVPQNGSTTLKLSPIYHYTRVTAINKYGGEPSPVSE---TYVTEPAPEK 716
QY 509 VDLDAVAGLEMMVNAEDVFPVMAVPMSP-----EHQLO-RLQVTRKLEMEEOA 556
Db 717 NPVDVKG-----EGNETNMVITMKPLRMMDNAPQVQRYVQMRQGTG--PMQOI 767
QY 557 A---ELVGSATPRLYLASNHSNKGHPRGRIOMLSFAGE-PLPONSNAKGSWERYQ 612
Db 768 VSDPFLVASTSTV-----PYELKQVAVNSOGGPEPO---VYIGYSGEYR 812
QY 613 LAVTORKEEPPSSSVFNQNDPAPATVDFSDFINNETIAGKDLVAMTAG-----FLH 665
Db 813 QAIPELGEIELNNSAVLVK--WRP-VDLAQVKH--LRGINVYVWEGSORKSKRHH 867
QY 666 IPHAEDIPNTVYVNGVGFRLRPNFDEDPSEFYSADSIYFRGDODA-----G 713
Db 868 KDHYVVPANTTSV---ILSGLRPSSYHLEVOAENG---RSGSPASEFTSTPEGVPG 919
QY 714 ACEVNPPLACLPQAAACAPDLPAFSGHG 740
Db 920 HPEALHLEQSNSTSLLRMOPPLSHNG 946

RESULT 6
US-09-314-242-2
; Sequence 2, Application US/09314242A
; Patent No. 6248575
; GENERAL INFORMATION:
; APPLICANT: Elizabeth J. Golightly
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
; TITLE OF INVENTION: Having L-Amino Acid Oxidase Activity
; FILE REFERENCE: 5556,200-US
; CURRENT APPLICATION NUMBER: US/09/314,242A
; EARLIER FILING DATE: 1999-05-18
; EARLIER APPLICATION NUMBER: 09/080,428
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Trichoderma harzianum
US-09-314-242-2
Query Match 2.5%; Score 100; DB 4; Length 617;

```

```

Best Local Similarity 23.9%; Pred. No. 0.3; Mismatches 83; Conservative 36; Mismatches 139; Indels 90; Gaps 18;
QY 44 LSRELTAVMRFLOTQRLG--FGLVDAQAAR-----PSDNCYSEYELQPPRAA 90
Db 18 LIREKVAKEUNILTERGEVPGIPPPNDEGRLEGGSYSHDNLPSPILXSI-----KRAL 71
QY 91 LAHLDRGSP--PPAREALAIYFGRQPOPNVSELVVG-----PLPSPYRQDYVERHG 142
Db 72 LKEAPRAEEELPPRK--VCJYAG-----VSGIYIMILDLKIPILTYDIFESSRTG 123
QY 143 GPLPHRRPVLFOEYLDIDOMIFENRELPOASGLLHCCFYKRGKRNLYTMATARGLOSG 202
Db 124 GRLTHHTDAKHDIYDIGAMRY-PDIPSK-----KFNPLKRTKMPILIKYYL 171
QY 203 DRATWGLYUYNISGAGFELHHVGLLELVN-----HKALDPAWTKQVYQGRYDS 254
Db 172 DGENTPOLYNN---HEFAKGVSDPYVAVSVANGTVPDDVDSVGEKLOQAF--GYKKEK 225
QY 255 LAOLEAGFAGLVNVLIPDNGTGSWSLSKSPVPGAPPLQFYPOGPRFSVQGSRAVS 314
Db 226 LAE---DEKGFDELMVDDMTT--REYLKRGKGPGEAPKXYD-----FAIQMETON- 273
QY 315 LMTFSGLGAFSGPRIFDVRFOGERLYEISLQ-----EALAIYGNS 357
Db 274 -----TGTNMLDQAF-SESVIDSFDFDNPTRPMTYCIGGTS 309

RESULT 7
US-09-063-733A-57
; Sequence 57, Application US/09063733A
; Patent No. 6372211
; GENERAL INFORMATION:
; APPLICANT: Isaac, Barbara G.
; APPLICANT: Greenplate, John T.
; APPLICANT: Purcell, John P.
; APPLICANT: Romano, Charles P.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROLLING
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: PO Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/063,733A
; FILING DATE: 21-Apr-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Patterson, Melinda L.
; REGISTRATION NUMBER: 33,062
; REFERENCE//DOCKET NUMBER: MOBT:022
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1400
; TELEFAX: 713-787-1440
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 617 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
US-09-063-733A-57
Query Match 2.4%; Score 95; DB 4; Length 617;
Best Local Similarity 22.5%; Pred. No. 1;

```

Matches 78: Conservative 40: Mismatches 142: Indels 86: Gaps 18:

QY 44 LSRELTAVMRELTORLG--PGLVDAQAAR-----PSDNCVSEVLOLPKAAA 90
18 LIREVAKELNLTRELGEVPGIPPPNEGRIFGGYSHDNLPDLSYSSI-----GGSGSG 73
QY 91 LAHLDRGSPPARREALATVFFGROPOPNVSELVVG-----PLPHSTYKRDVTVENHGR 144
Db 74 XXXXXXGGGPPRRK--VCYIAGAG-----VSGLYIAMILDLKIPNLTVDIFFSSSRRTGR 125
QY 145 LPHRRPVLFQEYLDIDOMIFNRELPOASGLHHCCFYKRGKRNLTMTTAPRGLOSGDR 204
Db 126 LTHHTDAKHDIYDIGAMRT-----PDIPSKRTFNLFRKTGAPL-----KYIYDGEN 175
QY 205 ATWFGLYYNISGAGFLHHVLELVN-----HKALDPARMTIQKVFYOGRYVDSLA 256
Db 176 TP--QLYNN-----HFEAKGVSDPVYVSVANGTVPDDVDVSGEKLOQAF--GYEKELKA 227
QY 257 QLEAOFEAGLVNVVLIIPDNGTGSNSLKSPPVPPGAPPLQFTPOGREFSVQGSRVASSLM 316
Db 228 E--DEDKGFDELMVDDMTT--REYLKRGKGPKEAPKYDF-----FAIOWMETON--- 273
QY 317 TFSFGAGFSPRIFFVROGERLVEISLQ-----ELALAYGNS 357
Db 274 -----TGTLNFDQAF--SESVIDSFDPDNFTKEWYICIBGTS 309

RESULT 8
US-09-136-574A-44
Sequence 44, Application US/09136574A
Patent No. 6294366
GENERAL INFORMATION:
APPLICANT: Farrington, Graham K.
Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.
TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
CITY: Spring House Corporate Center, P.O. Box 457
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136, 574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 1751 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-09-136-574A-44

Query Match 2.4% Score 94.5; DB 4; Length 1751:
Best Local Similarity 19.0% Pred. No. 6.9;
Matches 139; Conservative 83; Mismatches 284; Indels 225; Gaps 35;

QY 10 GGDGGEPSQLPHCPSPSPAPAP---THPGOSOLFDSLRELTAVMRELTORLGGLVD 66
Db 578 GGEVEHKKAQAKISV--PGYPMDFPNDPSYKGLTQLEKNKYIA----- 621
QY 67 AAQARPSDNCVTSVELQLPKAAALAHLDGRSPPARREALATVFFGROPOPNVSELVGP 126
Db 622 ---AYDNMNLVWGLE---PGAATSTPAPTSTPTPT-----PTPTPTVATPT-P 663
QY 127 LPHPS-----YMRDVTYERHGGPLPHRRPVLFQEYLDIDOM-----IFN 166
Db 664 TPTPTPTGSPGTGSGYKVLKNNETSASTGSIKPMKIVNGSSSVDLERKIRIYTYVD 723
QY 167 RELPOASGLHHCCFYKRGKRNLTMTTAPRGLOSGDRATWFGLYYNISGAGFLHHVGL 226
Db 724 GDNKQSA-----VCDMAQIGASNVTFNFK--LSSG-----VSGADVYL--EYGF 764
QY 227 ELLVNHKALDPARMT--IQKVFYOGRY-----YDSLQLEAOFEAG-----LVNVVLI 272
Db 765 SSGAGQ--LQPKGDTDIDQVRFKNDMSNYNQADWSMLQSMNYGENAKVTLTYVGVLV 822
QY 273 ----PDNGTGSWSLSKSPV-----PGAP-PLOFTYPOGREFSVQGSRVASSLMTFSF 320
Db 823 WGOEPGATPAPTSTPT 882
QY 321 GLGAFS-----GPRIFVROGERLVEI--SLOEALAYG 354
Db 883 SYGALKVWYANGNLSSPTVNLNKIKIENNGTAVDSRKVYVWTTIDOEALQSVVAS 942
QY 355 GNSPALMTTRYVDGFGMGKTYTPLTRGVDCPYLATIYVDMH-----LLESAPKTIK 407
Db 943 SINPAVIDVFKVLGANAG-----GAD--YVEIGFSGAGVLAAGOSTKEIRL 989
QY 408 ---DAFCVEQONQGLPRRHSDLSYHFGLAETVLYVMSMTLLNTYDVMTDTPVHPSG 464
Db 990 SIQSGGSINQSDYVRSANSTIENKTYGIDVL-----YKGR--EPGR 1034
QY 465 ALEIR-FYATGYISSAFLEFGATGKRYGNQVSEHTLGVTHSAHKYDLDAVAGLENNYMAE 523
Db 1035 NAOIKVWYANGNLGS-----MTVNLNPKIKIENNGT-----AVDSLKVKRYVWTTID 1082
QY 524 DMVFVPMVAPWSPDHOLQRLQVTRKLLMEMEQALVGSATPRYL--YLASHNSNKKRGP 561
Db 1083 GEA-----TOSVSYTSINPAVIDKFKVLGANNAGAD 1115
QY 582 RGRIGQLSFAGEPLPQNSSMARGFSWERYQVLAATQKKEEPSSSSVFNQNDPAPTVD 641
Db 1116 YVEIGFGSGAG-----VLAQOSTKEIRLSIQ-----GSGSINQSDYVRS 1160
QY 642 SDFINNETIAG 652
Db 1161 TGYIENKXVIG 1171

RESULT 9
US-08-471-119A-2
Sequence 2, Application US/08471119A
Patent No. 5827706
GENERAL INFORMATION:
APPLICANT: Lettner, Ernst
APPLICANT: Schneider, Elisabeth
APPLICANT: Schoergerdorfer, Kurt
APPLICANT: Weber, Gerhard
TITLE OF INVENTION: Cyclosporin Synthetase

```

NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 5827706artis Corporation
STREET: 59 Route 10
City: East Hanover
STATE: New Jersey
COUNTRY: USA
ZIP: 07936
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,119A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kassenoit, Melvyn
REGISTRATION NUMBER: 26,389
REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 503 8474
TELEFAX: 201 503 8807
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 15281 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHEetical: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Tolypocladium niveum
STRAIN: ATCC 34921
US-08-471-119A-2

Query Match
Best Local Similarity 2.4%; Score 94.5; DB 2; Length 15281;
Matches 154; Conservative 110; Mismatches 307; Indels 279; Gaps 41;

QY 26 SPSPQPTHTGSGOLPQADSLREELTAVMRLTORL-GPGLVDAQAAPSDNCVSEVLQ 84
DB 13682 NPDLATVDTSTKLTYALPQSDHVAAMLSKQKLPASIVYVLA PRSSSEIVACIGI-- 12739
QY 85 PPKAAALAHNDRGSPPARALAIIVFGGRQPPRVNSELVGPRLPHPSYMRDVIYERHGGP 144
DB 12740 --LKANLALYLPMDSNVEARQAAL--SETPEKFEVLGA-----GVP 12778
QY 145 LPYHRRFVLFOEYLDIDIMIFNRELPOASGLLHHCFCYKHHGRNLVTMTTAPRGLQSGDR 204
DB 12779 IPDKT-----ADV-RMVFISD-----IYASKTDKYSFGTR 12809
QY 205 ATMFGLYYNI--SGA-----GFLHVGLELL-----VNHKALDPARMT 241
DB 12810 PSASSLAYVIFTGSGTGRPGVMYEHGRVLSYKONASRIPOSRLMAHVSILAFDASVWE 12869
QY 242 IQGVFOG-----RY--DSLALQLEAOFEAGLVNVLIPDNTGGSWSLSKSPVPGP- 291
DB 12870 IFTTLNGGLFCISFTVLDKA--USAASDRHINITLLPRA-----LLKQCLADAPS 12922
QY 292 -----APPLQFPQGPFSYOG---SRVASSLW-----PFSFG--L 322
DB 12923 VLSLESLEYIGDRIDGADATKYKDLVKKAYNAVGTENSVMSTIYTIHEFTFANVPVI 12982
QY 323 GASSGR-----IDVNRQGERLYEISLOEALAIYGNSSPAAMTRRVYDGGFGMGKYYTP 378
DB 12983 GTSLGKSKAYIND--QDQOLV-----PAGVVGELVAVGDG----- 13016
QY 379 LFRGVCPYATYVDMHFLLESQAAPTIRDAFCVFEDNOGLPLRRHNSDLXSHFGSLAE 438
DB 13017 LANGYDPSLNTGRFTHITIDGQOVAYR-----TGDRVAYRPDRDYQIEFFRLDQ 13067

```

```

QY 439 TLVVRSMSTLNTANDYWDVYFHPSGAIEIRFATGYISSAFELGATGCKYGNV----- 492
DB 13068 QIKIRGHR-----TEPA-EVEQALLSDSSINAVVSAQNKEGLELVGYITT 13113
QY 493 -----SEHTLGTVHTHSAHF-----VLDVAGLE--NMVNAEDVFPMAVPW 534
DB 13114 QAAQSVKKEEKNVQOEKHAIFDSTAYANIGCIDRDALGDFLSTMSYDOSLIPR----- 13169
QY 535 SPEHOLQLOVT--RKLMEEQAAFL--VGSATPRYLYLASHNSKWKGPGRGRIQMLSFA 592
DB 13170 --EEMOEMLNDTMSRLDNPQPKVLEIGTGTVLF-----NLGVEG--LQSYA 13216
QY 593 GEPLPQSSMAKGSWERYQLAIVQKKEEPPSSSVYQNQDPAAPYDPSFINNETIA- 651
DB 13217 G--LEPSRSVT--AW-----VNKALETFFSLAGSARVHGTAEEDISSIDGLSLDLVI 13265
QY 652 -----GKDLVAVTAGEFLIHPAEDIPRTVTVGNGVGF--LRPY--NF----- 691
DB 13266 NSVAGYFPSPREYLAELTANLIRLPGVKRI-----FFGDMKRYATKDKLVARAV 13314
QY 692 -----FDEDPSTYSADSIYFRGDODAGACEVNPACLPOAAACAPDL 733
DB 13315 HTLGSNAKAMVROQVAKLEDEDEELLVDPAFETSLSDQFPDEIKHYELLPKRMAATNEL 13374
QY 734 PAFSHGFSH 743
DB 13375 SSRYAAVTH 13384

RESULT 10
PCT-US94-03437-8
Sequence 8, Application PC/TUS94/03437
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: HOMOGENEOUS IMMUNOASSAYS USING MUTANT-
NUMBER OF SEQUENCES: 124
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/03437
FILING DATE:
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 485 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEtical: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Leuconostoc dextranum
STRAIN: ATCC 19255
PCT-US94-03437-8

Query Match
Best Local Similarity 2.3%; Score 93.5; DB 5; Length 485;
Matches 81; Conservative 61; Mismatches 151; Indels 97; Gaps 20;

QY 387 YLATYVDMHF-----LLESQAAPTIRDAFCVFEDNOGLPLRRHNSDLXSHFGSLA 437
DB 30 YKGYIDGHEFAIVGTARQSLDDEFRQLVDSIKDFEDQA-----QAEAFIAHFYRA 83
QY 438 ETVLVYASMTL-----LNTDYWDVYFHPSGAIEIRFAT--GYISSAFELGATGK 487
DB 84 HDVTDAASYGILKSAIEAATKFDIDGNRIIFYMS--VAPRFEGTIAKYLKSEGLLAETG- 140

```

QY 488 YGNVSEHTLCTVHTSHAFKVDLDVAGLENNVW-----ADNVFVPMV----- 532
DB 141 YNRLMEKPFPGSYATAEELQSDLENAFDDDLFRIDHYLGKEMQONIALRFGNPDEDA 200
QY 533 PMSPEHOLQOVT-RKLEMEEOAFLVGSATPRYLASHNSKMGHPRGYRIQMLSF 591
DB 201 AMNKDY-1KNVQVTLAEVLGVEERAGY--DTTGALLDMIOQHT-----MQIVGM 247
QY 592 AGEPLQO--NSSMARGFSMERYOLAVTORKEEPSSSSVFNONDPWAPTVDSDFT---- 645
DB 248 LAMEKPEFNDKDIRAKNAAFN-ALKIYNEEVNKKYFVRAQYGA-GDTADYKPYLEAD 305
QY 646 -----NNETIAGK---DLVAVWTAGFLHHPHAEIPNTVTVGN-----GVGEFLR 687
DB 306 VPADSKNNTF1AGELQFDLPW-----EGVPYVRSGKRLAAKQTRVDIVFKAG 354
QY 688 PYNFDEDPFSYADSIFRGDODAGACEV 717
DB 355 TFFGSEQAOESVLSIT--DPKGAIEL 381

RESULT 11

US-07-737-071A-2
Sequence 2, Application US/07737071A
Patent No. 5229286
GENERAL INFORMATION:
APPLICANT: JARSCH, Michael
TITLE OF INVENTION: CLONING AND OVEREXPRESSION OF
TITLE OF INVENTION: GLUCOSE-6-PHOSPHATE DEHYDROGENASE FROM LEUCONOSTOC
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Armstrong, Nikolaio, Marmelstein Kubovcik &
STREET: 1725 K Street, N.W., Suite 1000
CITY: Washington D.C.
COUNTRY: United States of America
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/737,071A
FILING DATE: 19910730
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: DE P 4024158.9
FILING DATE: 30-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: Murray, Robert B.
REGISTRATION NUMBER: 22,890
REFERENCE/DOCKET NUMBER: 911028
TELEPHONE: (202) 659-2930
TELEFAX: (202) 887-0357
TELEX: 440142
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-737-071A-2

Query Match 2.3%; Score 93.5; DB 1; Length 486;
Best Local Similarity 20.8%; Pred. No. 0.96;
Matches 81; Conservative 61; Mismatches 151; Indels 97; Gaps 20;

QY 387 YLATYVDWMF-----LIESQAPKTIRDAFCVFEONQCLPLRRHSHLSIHYFGCLA 437

DB 31 YKGYLOEHFAIVGTARQOOLSDDEFKQOLVRSDINKDTEQA-----QAEFIAPHSYRA 84
QY 438 ETVLVVRSMSTL-----LNDYVWDTVFHPGAIETRYAT--GYISSAFLEGANGK 487
DB 85 HDVTDAA5YGLKSAIEEATATKFDIDGNRIFYMS--VAPREFGTAKYLSKLSGLAENG- 141
QY 488 YGNVSEHTLCTVHTSHAFKVDLDVAGLENNVW-----ADNVFVPMV----- 532
DB 142 YNRLMEKPFPGSYATAEELQSDLENAFDDDLFRIDHYLGKEMQONIALRFGNPDEDA 201
QY 533 PMSPEHOLQOVT-RKLEMEEOAFLVGSATPRYLASHNSKMGHPRGYRIQMLSF 591
DB 202 AMNKDY-1KNVQVTLAEVLGVEERAGY--DTTGALLDMIOQHT-----MQIVGM 248
QY 592 AGEPLQO--NSSMARGFSMERYOLAVTORKEEPSSSSVFNONDPWAPTVDSDFT---- 645
DB 249 LAMEKPEFNDKDIRAKNAAFN-ALKIYNEEVNKKYFVRAQYGA-GDTADYKPYLEAD 306
QY 646 -----NNETIAGK---DLVAVWTAGFLHHPHAEIPNTVTVGN-----GVGEFLR 687
DB 307 VPADSKNNTF1AGELQFDLPW-----EGVPYVRSGKRLAAKQTRVDIVFKAG 355
QY 688 PYNFDEDPFSYADSIFRGDODAGACEV 717
DB 356 TFFGSEQAOESVLSIT--DPKGAIEL 382

RESULT 12

US-08-022-096-2
Sequence 2, Application US/08022096
Patent No. 5308770
GENERAL INFORMATION:
APPLICANT: JARSCH, Michael
TITLE OF INVENTION: CLONING AND OVEREXPRESSION OF
TITLE OF INVENTION: GLUCOSE-6-PHOSPHATE DEHYDROGENASE FROM LEUCONOSTOC
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nikolaio, Marmelstein, Murray & Oram
STREET: 655 Fifteenth Street N.W. Suite 330
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-5701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/022,096
FILING DATE: 19930225
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: DE P 40 24 158.0
FILING DATE: 30-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: Chin, Monica F.
REGISTRATION NUMBER: 36,105
REFERENCE/DOCKET NUMBER: P564-3005
TELEPHONE: (202)638-5000
TELEFAX: (202)638-4810
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/445,464C
APPLICATION NUMBER: US/08/445,464C
FILING DATE: 22-MAY-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/044,857
FILING DATE: 08-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Ruzsala, Lois K.
REGISTRATION NUMBER: 39,074
REFERENCE/DOCKET NUMBER: BEH-7261 DIV 1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (847) 267-5364
TELEFAX: (847) 267-5376
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-445-464C-8

Query Match 2.3%; Score 93.5; DB 3; Length 486;
Best Local Similarity 20.8%; Pred. No. 0.96;
Matches 81; Conservative 61; Mismatches 151; Indels 97; Gaps 20;

387 YLATYVDMHF-----LLESOAPRTIRDAFCVFONOGDLPLRRHSDLSHYFGGLA 437
31 YKKGYLQEHRAIVGTARQQLSDDEFKQLVRSIDKFEDDA-----QAEAFIHFYSRA 84
438 ETVLVVSMSTL-----LNTDYVMDVTFHPSGAIEIRFYAT--GYISSAFLEGATGK 487
85 HDVTDAASYGILKSAIEAATKFDIDGNRIFYMS--VAPRFPGTIATYKSEGGLAETG- 141
488 YGNQVSEHTLGTVHTSAHKVVDLVAGLENMVA-----AEDVFTYPMAY----- 532
142 YNRLMIKPEFGTSYATAEELQSDLENAFDDQLFRIDHYLQKEKVNQNIALRFGNPIFDA 201
533 PMSPEHOLQLOVT--RKLEMEBOAFLVGSATPRYLYLASNHSNKGHPRGYRIOMLSF 591
202 AMNKDY--IKNQVTLAEVLGEERAGY--DTTCALLDMIONHT-----MQYGVW 248
592 AGEPLQ--NSSMARGESWERYQLAVTQRKEEPPSSSVFNQNDPMAPTVDSDFI---- 645
249 LAMEKPEPFNDKDIRAKNAF--ALKIYNEEVNKYFVRAQYGA--GDTADYKPYLEAD 306
646 -----NNETIAGK--DLVAVTAGFLIHPIHADIPTVTYGN-----GVGFPLR 687
307 VPADSKNNTFIAGELQFLDPRM-----EGVPPYVNSGRLAKAKOTRVDIVFKAG 355
688 PYNFDEDPFSYSDSIYFRGDODAGACEV 717
356 TFMGSEDAQESVLSITII---DPKGAIEL 382

RESULT 15
US-08-044-857D-8
Sequence 8, Application US/08044857D
Patent No. 6455288
GENERAL INFORMATION:
APPLICANT: Jakobovits, Edward B.
Silen, Joy L.
Levy, Mark J.
Goodman, Thomas C.
Becker, Martin

Caldwell, Robert M.
Bolt, Richard R.
Barnett, Christopher C.
TITLE OF INVENTION: Homogeneous Immunoassays Using Mutant Glucose-6-Phosphate Dehydrogenases
NUMBER OF SEQUENCES: 124
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dade Behring Inc.
STREET: 1717 Deerfield Road
CITY: Deerfield
STATE: Illinois
COUNTRY: USA
ZIP: 60015

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/044,857D
APPLICATION NUMBER: US/08/044,857D
FILING DATE: 08-APR-1993
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Ruzsala, Lois K.
REGISTRATION NUMBER: 39,074
REFERENCE/DOCKET NUMBER: BEH-7261
TELECOMMUNICATION INFORMATION:
TELEPHONE: (847) 267-5364
TELEFAX: (847) 267-5376
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Leuconostoc dextranlicum
STRAIN: ATCC 19255
SEQUENCE DESCRIPTION: SEQ ID NO: 8
US-08-044-857D-8

Query Match 2.3%; Score 93.5; DB 4; Length 486;
Best Local Similarity 20.8%; Pred. No. 0.96;
Matches 81; Conservative 61; Mismatches 151; Indels 97; Gaps 20;

387 YLATYVDMHF-----LLESOAPRTIRDAFCVFONOGDLPLRRHSDLSHYFGGLA 437
31 YKKGYLQEHRAIVGTARQQLSDDEFKQLVRSIDKFEDDA-----QAEAFIHFYSRA 84
438 ETVLVVSMSTL-----LNTDYVMDVTFHPSGAIEIRFYAT--GYISSAFLEGATGK 487
85 HDVTDAASYGILKSAIEAATKFDIDGNRIFYMS--VAPRFPGTIATYKSEGGLAETG- 141
488 YGNQVSEHTLGTVHTSAHKVVDLVAGLENMVA-----AEDVFTYPMAY----- 532
142 YNRLMIKPEFGTSYATAEELQSDLENAFDDQLFRIDHYLQKEKVNQNIALRFGNPIFDA 201
533 PMSPEHOLQLOVT--RKLEMEBOAFLVGSATPRYLYLASNHSNKGHPRGYRIOMLSF 591
202 AMNKDY--IKNQVTLAEVLGEERAGY--DTTCALLDMIONHT-----MQYGVW 248
592 AGEPLQ--NSSMARGESWERYQLAVTQRKEEPPSSSVFNQNDPMAPTVDSDFI---- 645
249 LAMEKPEPFNDKDIRAKNAF--ALKIYNEEVNKYFVRAQYGA--GDTADYKPYLEAD 306
646 -----NNETIAGK--DLVAVTAGFLIHPIHADIPTVTYGN-----GVGFPLR 687
307 VPADSKNNTFIAGELQFLDPRM-----EGVPPYVNSGRLAKAKOTRVDIVFKAG 355
688 PYNFDEDPFSYSDSIYFRGDODAGACEV 717

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 12:04:34 ; Search time 25 Seconds
(without alignments)
2951.379 Million cell updates/sec

Title: US-10-081-408-2_COPY_20_763

Perfect score: 4010
Sequence: 1 LVCVLLVGRGGDGEPSQLP.....QAAACAPDLPAFSGHGSFHN 744

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PTCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PTCT_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US50_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US50_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4010	100.0	763	US-10-081-408-2	Sequence 2, Appl1
2	3967	98.9	998	US-10-081-408-20	Sequence 20, Appl1
3	2489.5	62.1	729	US-09-919-497-51	Sequence 51, Appl1
4	106	2.6	617	US-10-005-530-58	Sequence 58, Appl1
5	104	2.6	980	US-09-908-193-22	Sequence 22, Appl1
6	103	2.6	2224	US-10-115-563-14	Sequence 14, Appl1
7	102	2.5	1194	US-09-738-626-5832	Sequence 5832, Ap
8	101	2.5	617	US-10-005-530-46	Sequence 46, Appl1
9	99	2.5	944	US-10-213-990-27	Sequence 27, Appl1
10	95.5	2.4	581	US-09-738-626-4914	Sequence 4914, Ap
11	95	2.4	617	US-10-005-530-57	Sequence 57, Appl1
12	93.5	2.3	878	US-09-912-020-364	Sequence 364, App
13	93	2.3	601	US-09-925-301-844	Sequence 844, App
14	92.5	2.3	635	US-10-009-895-33	Sequence 33, Appl1
15	92	2.3	904	US-09-712-363-247	Sequence 247, Appl1
16	91.5	2.3	390	US-10-004-551-16	Sequence 16, Appl1
17	91	2.3	990	US-09-738-626-6558	Sequence 6558, Ap
18	91	2.3	1247	US-09-808-193-2	Sequence 2, Appl1
19	91	2.3	2507	US-09-819-104A-2	Sequence 2, Appl1

20	90.5	2.3	388	US-09-361-630-1	Sequence 1, Appl1
21	90	2.2	589	US-09-740-041-2	Sequence 2, Appl1
22	90	2.2	617	US-10-005-530-56	Sequence 56, Appl1
23	89.5	2.2	630	US-10-101-464A-71	Sequence 71, Appl1
24	89.5	2.2	911	US-09-855-754-4	Sequence 4, Appl1
25	89.5	2.2	911	US-10-227-353-4	Sequence 4, Appl1
26	89.5	2.2	953	US-09-845-583-4	Sequence 4, Appl1
27	89.5	2.2	3695	US-10-037-182-2	Sequence 2, Appl1
28	89	2.2	365	US-09-925-299-1007	Sequence 1007, Ap
29	89	2.2	365	US-09-925-299-1007	Sequence 1007, Ap
30	89	2.2	1430	US-10-087-993-34	Sequence 34, Appl1
31	88.5	2.2	583	US-09-942-185-2	Sequence 2, Appl1
32	88.5	2.2	616	US-09-925-300-1519	Sequence 1519, Ap
33	88.5	2.2	882	US-10-174-590-574	Sequence 574, App
34	88.5	2.2	882	US-10-176-758-574	Sequence 574, App
35	88.5	2.2	882	US-10-175-737-574	Sequence 574, App
36	88.5	2.2	882	US-10-173-706-574	Sequence 574, App
37	88.5	2.2	882	US-10-173-738-574	Sequence 574, App
38	88.5	2.2	882	US-10-175-752-574	Sequence 574, App
39	88.5	2.2	882	US-10-176-482-574	Sequence 574, App
40	88.5	2.2	882	US-10-176-757-574	Sequence 574, App
41	88.5	2.2	882	US-10-176-913-574	Sequence 574, App
42	88.5	2.2	882	US-10-180-552-574	Sequence 574, App
43	88.5	2.2	882	US-10-180-557-574	Sequence 574, App
44	88.5	2.2	882	US-10-173-700-574	Sequence 574, App
45	88.5	2.2	882	US-10-174-572-574	Sequence 574, App

ALIGNMENTS

RESULT 1
US-10-081-408-2
Sequence 2, Application US/10081408
Patent No. US20020160482A1
GENERAL INFORMATION:
APPLICANT: Abirams, n, Lars
TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
FILE REFERENCE: 13425-053001
CURRENT APPLICATION NUMBER: US/10/081,408
CURRENT FILING DATE: 2002-02-21
PRIOR APPLICATION NUMBER: SE 0100625-3
PRIOR FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: US 60/272,247
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 20
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 2
LENGTH: 763
TYPE: PRT
ORGANISM: Homo sapiens
US-10-081-408-2

Query Match 100.0%; Score 4010; DB 9; Length 763;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 744; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LVCVLLVGRGGDGEPSQLPHCPSPSPAPWTHPQOSQFALDSREELTAVNRFTLRL 60
DB 20 LVCVLLVGRGGDGEPSQLPHCPSPSPAPWTHPQOSQFALDSREELTAVNRFTLRL 79
QY 61 GPELVAAQARPSPDNCFVSVELDLPKKAALALHLDGSPPARALATVFFGQPOPWNS 120
DB 80 GPELVAAQARPSPDNCFVSVELDLPKKAALALHLDGSPPARALATVFFGQPOPWNS 139
QY 121 ELTVGGLPHPSYRDVTVRRHGPPLPYHRRPVLFQEYLDIDOMIFNRELPOASGLHHCC 180
DB 140 ELTVGGLPHPSYRDVTVRRHGPPLPYHRRPVLFQEYLDIDOMIFNRELPOASGLHHCC 199
QY 181 FYKHRRNLVTMTATARGLOSGRATWFGLYNYISGAGFLHVGLELLVNHKALDPARW 240
DB 200 FYKHRRNLVTMTATARGLOSGRATWFGLYNYISGAGFLHVGLELLVNHKALDPARW 259

```

OY 241 TIOKVFQGRYVDSLADLEAOFEGAGLVNVTIIPONGTGSNLSKSPVPGCAPPLQLOFPQ 300
Db 260 TIOKVFQGRYVDSLADLEAOFEGAGLVNVTIIPONGTGSNLSKSPVPGCAPPLQLOFPQ 319
OY 301 GPRFSVGSRASSLMTFSEFGLGAFSGGRIFDVFQGERLYEISLQBALIYGGNSPAA 360
Db 320 GPRFSVGSRASSLMTFSEFGLGAFSGGRIFDVFQGERLYEISLQBALIYGGNSPAA 379
OY 361 MTRRYVDGFGMGKYYTPPLRGVDCPYLATIYVDNHFLLESQAPKTIIDAFCVFEQNOGLP 420
Db 380 MTRRYVDGFGMGKYYTPPLRGVDCPYLATIYVDNHFLLESQAPKTIIDAFCVFEQNOGLP 439
OY 421 LRRHNSDLXSHYFEGGLAEVTVLVVBSMSTLLNXYDVWQVHFPPSCAIEIRFATGYSASF 480
Db 440 LRRHNSDLXSHYFEGGLAEVTVLVVBSMSTLLNXYDVWQVHFPPSCAIEIRFATGYSASF 499
OY 481 LFGATGKXGNVDSHETLGTVHTHSAHFYVDLDVAGLENNWMAEDVAFYPMAYPMSPEQOL 540
Db 500 LFGATGKXGNVDSHETLGTVHTHSAHFYVDLDVAGLENNWMAEDVAFYPMAYPMSPEQOL 559
OY 541 QRLQVTRKLLMEBOQAEVLVGSATPRYLYLASNHSNKGHPRGYRIQMLSFAGEPLPONS 600
Db 560 QRLQVTRKLLMEBOQAEVLVGSATPRYLYLASNHSNKGHPRGYRIQMLSFAGEPLPONS 619
OY 601 SMAGFSMERQOLAVTQKKEEPPSSSYFNQNDPAPVYVDSDFINNTITAGKDLVAVYT 660
Db 620 SMAGFSMERQOLAVTQKKEEPPSSSYFNQNDPAPVYVDSDFINNTITAGKDLVAVYT 679
OY 661 AGFLIIPAEIDIPMTVTYVGNVGFFLRPNYFNFEDEPSFYSDSIYFRGDODAGACEVNP 720
Db 680 AGFLIIPAEIDIPMTVTYVGNVGFFLRPNYFNFEDEPSFYSDSIYFRGDODAGACEVNP 739
OY 721 ACLPQAAAACAPDLPAFSHGFSHN 744
Db 740 ACLPQAAAACAPDLPAFSHGFSHN 763

RESULT 2
US-10-081-408-20
; Sequence 20, Application US/10081408
; Patent No. US20020160482A1
; GENERAL INFORMATION:
; APPLICANT: Abrams,n , Lars
; APPLICANT: Nilsson, Joakim
; TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
; FILE REFERENCE: 13425-053001
; CURRENT APPLICATION NUMBER: US/10/081,408
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: SE 0100625-3
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 60/272,247
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 998
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant construct
US-10-081-408-20

```

	Query Match	98.9%	Score 3967:	DB 9,	Length 998;
	Best Local Similarity	99.9%	Pred. No. 0;		
	Matches 735:	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	9 RCGDGGESQLHCHCSVSPSAQPMTHPCOSQALDSREELTAVMRFLTRLQGLGVDA	68			
Dd	263 QGDDGGESQLHCHCSVSPSAQPMTHPCOSQALDSREELTAVMRFLTRLQGLGVDA	322			
Oy	69 QARPSDNCFYSVELDLPKAAALAHLDKSGPPAREALAIYFFGNQPDPNVSSELVGPPLP	128			

Db	323	QARSNDNCVFSVEQLRPPKAAALAHDRGSPPRAREMLAIVFFGRQRPVSEVLVYGP	382
QY	129	HPSYRBDVYERHGGPLPYHRRPVLFQEYLDIDQIMFNRELPOASGLLHHCCFYKHKGRN	188
Db	383	HPSYRBDVYERHGGPLPYHRRPVLFQEYLDIDQIMFNRELPOASGLLHHCCFYKHKGRN	442
QY	189	LYMTTTPARGLQSSDRATMPCFLYNNISGAPFELHHVGLLELVNKKALDPAKMTIOKFYQ	248
Db	443	LYMTTTPARGLQSSDRATMPCFLYNNISGAPFELHHVGLLELVNKKALDPAKMTIOKFYQ	502
QY	249	GRYVDSLQALQAEQAEGLVNVNVLIPDNGTGGSMELKSPVPPGAPRLQFYPOGPRFSVQ	308
Db	503	GRYVDSLQALQAEQAEGLVNVNVLIPDNGTGGSMELKSPVPPGAPRLQFYPOGPRFSVQ	562
QY	309	SRVASSLMTFSFGIGATSPGRIEDVRROGERLYVEISLOELALYGGNSPAAMTTRVYDG	368
Db	563	SRVASSLMTFSFGIGATSPGRIEDVRROGERLYVEISLOELALYGGNSPAAMTTRVYDG	622
QY	369	GFGMGKTYTPTLRGVDCPYLATVYDMHFLLESQAPKTRDAFCVFEONOGCLPLRHHSDL	428
Db	623	GFGMGKTYTPTLRGVDCPYLATVYDMHFLLESQAPKTRIDAFVCFEONOGCLPLRHHSDL	682
QY	429	YSHRFGGLAEVLYVYRSMSTILNTDYWDVYFHRSGALIEIRFYATGIISSAFLEGAIGKY	488
Db	683	YSHRFGGLAEVLYVYRSMSTILNTDYWDVYFHRSGALIEIRFYATGIISSAFLEGAIGKY	742
QY	489	GNQVSEHTLGVTHHSAPFKVDLDVAGLENNVMAEDVYFVMAVPASPEHQRLQYTRK	548
Db	743	GNQVSEHTLGVTHHSAPFKVDLDVAGLENNVMAEDVYFVMAVPASPEHQRLQYTRK	802
QY	549	LLEMEQAAFLVGSATPRYLYLASNHSKKMGHPGQYRIOMISFGGEPLPONSSVARGFSW	608
Db	803	LLEMEQAAFLVGSATPRYLYLASNHSKKMGHPGQYRIOMISFGGEPLPONSSVARGFSW	862
QY	609	ERYOLAATVORKEEPPSSSVNONDPMAPTVDFSDFINNETIACKDLVAVYTAGFLHIPH	668
Db	863	ERYOLAATVORKEEPPSSSVNONDPMAPTVDFSDFINNETIACKDLVAVYTAGFLHIPH	922
QY	669	AEDIPNTVYVNGVGFELRPYNFDEDESPYSADSITFRGDQDAGACEVNPDLACLPAAA	728
Db	923	AEDIPNTVYVNGVGFELRPYNFDEDESPYSADSITFRGDQDAGACEVNPDLACLPAAA	982
QY	729	CAPDLPARFSGHGSFHN 744	
Db	983	CAPDLPARFSGHGSFHN 998	
RESULT 3			
US-09-919-497-51			
: Sequence 51, Application US/09919497			
: Patent No. US2002010662A1			
: GENERAL INFORMATION:			
: APPLICANT: Muller, George L.			
: TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER			
: FILE REFERENCE: BOB01/7725			
: CURRENT APPLICATION NUMBER: US/09-919,497			
: CURRENT FILING DATE: 2001-07-31			
: PRIOR APPLICATION NUMBER: US 60/221,735			
: PRIOR FILING DATE: 2000-07-31			
: NUMBER OF SEQ ID NOS: 100			
: SOFTWARE: PatentIn version 3.0			
: SEQ ID NO 51			
: LENGTH: 729			
: TYPE: PRT			
: ORGANISM: Homo sapiens			
US-09-919-497-51			

[illegible]

```
Db 20 LAYVHLLTSPGSGS---SOPPHCPVSHRAQPMHPHGOQLFADLSHEELTAVMRLTQRL 75
Oy 61 GGLVDAQAARPSDNCVFSEVLEQLPPKAAALAHLDGSPPPAREALATVFGQPOPNIS 120
Db 76 GGLVDAQAARPSDNCVFSEVLEQLPPKAAALAHLDGSPPPAREALATVFGQPOPNIS 135
Oy 121 ELVVGRLPHPSYMRDVTVERHGGPLRYHRRPYLFQELIDIDIMIFNRELPOASGLLHCC 180
Db 136 ELVVGRLPHPSYMRDVTVERHGGPLRYHRRPYLRAEFTQMRRHLKQVLEPKAPILFSSR- 194
Oy 181 FKHGRNLVMTAPRGQSGDRATWFGLYTISGAGFFLHVGLLELVNKHADLPARM 240
Db 195 -FNYNSTLAAVAHATPRGRSRRRTWIGLYNHSIGVGLFHPVGLLELDHDLADPAM 253
Oy 241 TIOKVGRCYVDSLAQLOAFGLVNVLLIPDNGTGSWSLKSVPVPPAPLPFYQ 300
Db 254 TVQGVFLGHYYADLQGLERKSGRLVRYVLPPLPPNGASLSRNSFGPLPLPFSQ 313
Oy 301 GPRFVSQGRVASSLMTFSGGLAFSGPRIFDVRFQGERLYVEISLQELALTYGNSPAA 360
Db 314 GSQSVQGNLVVSSLSMFTFSGHGVSGRLFDVRFQGERLAYEVSVQECVSIYGADSPKT 373
Oy 361 MTRIVYDGGFGMKITTPRLRGYDCRYLATYVDMHFLLESQAKTRDAFCVEQNGLP 420
Db 374 MLTRYLDSFGGLRNSRGVGYDQYQATMDIHILVKGAVQLPFGACVFEEOGLP 433
Oy 421 LRRHSDLSHYEGGLAETVLYVRSMTLLNDYVMDTVFHPGSALEIRFATGYISSAF 480
Db 434 LRHHNYLQNHFEYGLASSALVYRSVSVGNVYIMDEVLYPBGALGRHAHYINTAF 493
Oy 481 LEGATG--KYGNVSEHTLGTVTHSAHFKVDLDVAGLNMVAEDMVPVMAVPMSPH 538
Db 494 LKGEGLLFGNRYGVRGTVTHAFHEKLDLDVAGLKMVVAEDVFEKPVAA PNNPEH 553
Oy 539 QLORLVTRKLEMEBOAFLVGSATPRYLYLASNHNKMGHGRYRIQMLSTAGEPLPQ 598
Db 554 WLORPOLTRVYLGKEDITAFSLGSPRLRYLYLASNOTNAMGHORG----- 598
Oy 599 NSSMAGFSEWERYQALVTRKKEEPPSSSVFNQNDPAPVDESDFINNETIAGKDLVAM 658
Db 599 -----YQLVYTRQKKEEESQSSSYHQNDLWTPTVTADEFINNETILLGEDLVAM 646
Oy 659 VTAGFLHIAHEDIPNTVYVANGVGFLLRYNFDDEDPSTYSADSIYFRDQAGACEVN 718
Db 647 VTASFLHIAHEDIPNTVYVANGVGFLLRYNFDDEDPSTYSFSGSVYFEKGODAGLCSIN 706
Oy 719 PLACLPQAAACAPDLPAFSGGF 741
Db 707 PVACLPLQAAACVPLDLPFSTHGF 729
```

```
RESULT 4
US-10-005-530-58
; Sequence 58, Application US/10005530
; Publication No. US20030026795A1
; GENERAL INFORMATION:
; APPLICANT: Isaac, Barbara G.
; APPLICANT: Greenplate, John T.
; APPLICANT: Purcell, John P.
; APPLICANT: Romano, Charles P.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROLLING INSECTS
; FILE REFERENCE: 11899.0022.DVUS01 (MOBT:022--2)
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US/10/005,530
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/044,504
; PRIOR FILING DATE: 1997-04-21
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58
; LENGTH: 617
; TYPE: PRT
```

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Polypeptide
; NAME/KEY: MISC_FEATURE
; LOCATION: (74)..(79)
; OTHER INFORMATION: Xaa = Unknown
US-10-005-530-58
Query Match 2.6%; Score 106; DB 9; Length 617;
Best Local Similarity 22.8%; Pred. No. 1.3;
Matches 79; Conservative 40; Mismatches 141; Indels 86; Gaps 18;
```

```
Oy 44 LSREELTAVMRLTQRLG--PGLVDAQAAR-----PSDNCVFSEVLEQLPPKAA 90
Db 18 LIREKVAKEILNTERLGEVGPFPNNEGRLFGGYSNDLPSDPLYSIK-----PGCG 73
Oy 91 LAHLDSGPPAREALATVFGQPOPNISLVYG-----PLPHPSYMRDVTVERHGGP 144
Db 74 XXXXXGGGPPRK--VCITGAG-----VGLYIAMLIDLKIPNLTYDIFESSSTTGR 125
Oy 145 LPYHRRPVLFOEYLDIDIMIFNRELPOASGLHCCFFYKHGRNLVMTAPRGQSGDR 204
Db 126 LYTHHTDAKHQDYDGNARY-----PDIPSMKRTFLFKRTGMLI-----KYLDGEN 175
Oy 205 ATWFGLYYINSGAGFFLHVGLLELVN-----HKALDPARMTOQVYQGRYVDSLA 256
Db 176 TP--QLYNN---HFEAKGVSDPYNVSANGTVDPDQVDSVGEKLOQAF--GYREKLA 227
Oy 257 QLEAOFAGLVNVLLIPDNGTGSWSLKSVPVPPAPLPQFYQGRFVSQGRVASSLSM 316
Db 228 E--DFDKGDELMLVDMDMT--REYLKRGGRGAEAPKDF-----FALQMETON--- 273
Oy 317 TFSFGGLGAFSGPRIFDVRFQGERLYVEISLQ-----EALAIYGNIS 357
Db 274 -----TGTNLFPQAF--SESVIDSFDEPDNPTKREWCIEGCTS 309
```

```
RESULT 5
US-09-908-193-22
; Sequence 22, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: SHIMKETS, RICHARD A.
; APPLICANT: ZERHUSEN, BRYAN
; APPLICANT: MALYANKAR, URIEL M.
; APPLICANT: PADIGARU, MURALIDHARA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 21402-062
; CURRENT APPLICATION NUMBER: US/09/908,193
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: 60/220,273
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: 60/221,650
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/221,233
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/220,912
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/218,875
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,870
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,901
; PRIOR FILING DATE: 2000-07-18
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 980
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-908-193-22
```

Query Match 2.6%; Score 104; DB 9; Length 980;
Best Local Similarity 20.0%; Pred. No. 3.9;
Matches 13; Conservative 69; Mismatches 204; Indels 262; Gaps 37;

```

214 ISGAGFLHVGLELVNHNKALDPA---WTIOKVYQGRYDSLAQLE-AQEGAG-LV 267
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
388 ISGKYLWREVGAEENANGRLPGRGDQAMDVGPV---RLKKKVKQYELTQLVPGRL 443
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
268 NVYLI-P---DNGGSGMSLSPVPBPAPLPLOFYPOGPRRSYOG-SRVASLTFTSGL 322
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
444 EVKLVAENKHEEDGYAAVWKGKTEKAPADMDIQGPPLPAHVAESSTSTIV-LRMKK 502
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
323 GAFSGPRI-FDVR-----OGERLYEISLQELAIYGNSPAMATTRY----- 365
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
503 PDFETVAIVNTVAFSPMGLRNASLVYTTSSGEDILI-GGLKP---FTKFEFAVQSHGV 558
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
366 -VDCGFPM-----GKTYPLTRGVDCPYLATVYDMHFLLESQAPKTIIDACVEON 416
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
559 DMDGPFSGSVYERSTLPDRSPSTPDLRLSPSTVRLHMCPTPEPNCIEIYELIYSSN 618
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
417 OGFLRR-----HHSDLYSHY-----FGSLATVIVRSM 447
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
619 HTQERHMTLLTTCGNIFSAEVLHESDTRIFFKMGARTEVGPFPFSLQDVIITLQEKLS 678
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
448 TLNLYDVMWDTVFHPSGAIEIRFYATGYISSAFL-----FG 483
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
679 DSLMHNV-----TGIVGVCLGLCLLACMCAGLRSPHRESLPELS 721
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
484 ATGVYGNV-SEHTLCTVHTSHAFKVDLDVAGLENNVMAEDNVFPMVAVPMSPEHOLQ 541
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
722 STAPGPALYSRARLGPSPPAH-----ELESIVH-----PHODWSPPPS- 764
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
542 RLQYTRKLEMEQAAF-LVGSATPRVLYLASNHSK---WGHPRGVRIMLSFAPGRL 596
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
765 -----DVEDRAEVHSLMGGVSE-----GRSHSKKKSMAQSG-----LSMAG-- 803
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
597 PONSMAFGSEMYEYOL-----AVTQ-----RKEEEP 623
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
804 -----SMACCELPAQAPRALTRALLPAGTGTLLQALVYDAIKNGRKKRPP 853
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
624 SSSVFNONDPAPVTPVPSDFINNETIAGKDLVAMVTAAGFLIHAEEDI-----PNTV 676
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
854 AC-----RNOVEAEVIVHSDP--SASNGNPDL-----HLODLEBEDPLPPBPAP 894
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
677 TVNGVGFELRPVDFPDPSFYASDSIYFRG---DODAGACEY-----NPLACLPOAA 727
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
895 DLISGV-----DPC-----OGAAMLDRELGCCELAPGPRLTCLPEAA 934
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
728 ACA---PDL 733
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
935 SASCSYPLD 943
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 6
US-10-115-563-14

; Sequence 14, Application US/10115563
; Publication No. US20030008307A1

GENERAL INFORMATION:

APPLICANT: Griffith, John H

Greenard, Judith S

TITLE OF INVENTION: METHODS FOR DIAGNOSING ACTIVATED PROTEIN

C RESISTANCE ASSOCIATED WITH A FACTOR V GENETIC MUTATION

AND COMPOSITIONS THEREOF

NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESS:

ADDRESSEE: The Scripps Research Institute, Office of

Patent Counsel

STREET: 10666 No. US20030008307A1th Torrey Pines Road, TPC 8

CITY: La Jolla

STATE: CA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/115,563

FILING DATE: 02-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/410,488

FILING DATE: 24-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Fitting, Thomas

REGISTRATION NUMBER: 34,163

REFERENCE/DOCKET NUMBER: 449.0

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-554-2937

TELEFAX: 619-554-6312

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 2224 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 14:

US-10-115-563-14

Query Match 2.6%; Score 103; DB 9; Length 2224;
Best Local Similarity 20.7%; Pred. No. 16;
Matches 76; Conservative 46; Mismatches 123; Indels 122; Gaps 20;

```

150 RPLVFOEYLDIDOMIFNRELPOA--SGLLHHCFF-----YKHRGRNLVTMTAPR 197
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
62 KKIYREY-----EYFKKFKQSTISGLGPTLYAEVGDIIKVFKNKADK-PLSIHPQ 115
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
198 GLQSGDRATWFGLYYNTS-GAGFELHVGLELVNHNKALDPA---WTIOK----- 244
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
116 GIR-----YSKLSSEGA SYLDHTFPARKM--DDAVAGREYTYEWSISEDSGPTHD 163
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
245 -----VFYGRYDLSLAQLEAFEGLVNVVLIIPNGT---CG----- 279
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
164 DPCLTHLY-----SHENLIEDFNSGLIGLIPDLICKKGLTEGTOKTEKQIYLLFAV 217
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
280 -----SMLKSPVPBPAPLPLOFYPOG--PRFSYOGSRVASSLTWTFSGLAGFSGPRI 332
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
218 FDESKSMQSSSL---MYTVNGYVNGTMDPTIYCAHDHS--WHL---LQMSGPELFS 268
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
333 VRQGERLYEISLQELAIYGNSPAMATTRYVDGFGMGKYYTP-LTRGVDCPYLATY 391
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
269 IHFNGQVLEQNHKHYSAITLVASATSTANMTVGPEKWIISLTPKHLQAG-----MQAY 323
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
392 VD-----WHFLLESQAPKTIIDACVCFEQNGCLPLRRHSD 427
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
324 IDINCPKRTNLRKLTREQRHMKRWEYFL--AAEVIWDYADVIPANMDKTYRSOHL 381
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
428 LYSHYFG 434
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
382 NFSNIG 388
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 7
US-09-738-626-5832

; Sequence 5832, Application US/09738626

; Publication No. US20020197605A1

GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI

APPLICANT: MIZOGUCHI, HIROSHI

APPLICANT: ANDO, SEIKO

APPLICANT: HAYASHI, MIKIRO

APPLICANT: OCHIAI, KEIKO

APPLICANT: YOKOI, HARUHIKO

APPLICANT: TATEISHI, NAOKO

APPLICANT: SENOH, AKIHIRO

Page 5

Query Match	2.5%;	Score 102;	DB 9;	Length 1194;
Best Local Similarity	18.9%;	Pred. No. 7.9;		
Matches 112;	Conservative 71;	Mismatches 207;	Indels 204;	Gaps 29;

RESULT 8
US-10-005-530-46
Sequence 46, Application US/10005530
Publication No. US20030026795A1
GENERAL INFORMATION
APPLICANT: Isaac, Barbara G.
APPLICANT: Greenplate, John T.
APPLICANT: Purecell, John P.

Query Match 2.5%; Score 101; DB 9; Length 617;
Best Local Similarity 23.3%; Pred. No. 3.6;
Matches 81; Conservative 39; Mismatches 138; Indels 90; Gaps 19;

RESULTS
US-10-213-990-27
: Sequence 27, Application US/10213990
: Publication No. US20030082595A1
: JOURNAL: INFORMATION

Query Match	2.58;	Score 99;	DB 9;	Length 944;
Best Local Similarity	20.28;	Pred. No. 10;		
Matches 122;	Conservative 62;	Mismatches 243;	Indels 176;	Gaps 29

LOCATION: (74)..(79)
OTHER INFORMATION: xaa - unknown
US-10-005-530-57

Query Match 2.4%; Score 95; DB 9; Length 617;
Best Local Similarity 22.5%; Pred. No. 12;
Matches 78; Conservative 40; Mismatches 142; Indels 86; Gaps 18;

QY 44 LSRELTAVMRRLTORLG--PGLVDAQAAR-----PSDNCVFSVELQLPCKAA 90
DB 18 LIREVAKELNITLERLGEVPIPPNEGRTFGGSHNDLPSDPLYSI-----GGSGG 73
QY 91 LAHLDRGSPPARREALATYFFGRQROPANSELVYG-----PLPHSYKRDVTVEHAGCP 144
DB 74 XXXXXXGGGPPRRK--VCYAGAG-----VSGLYIAMILDOLKIPNLTYLIFESSSTGR 125
QY 145 LPHYHRPVLFOEYLDIDQIFNRELPOAGSLHCCFYKRGHNLVTMTAPRGLQSGDR 204
DB 126 LYTHHTDAKHXYDYGAMRY-----PDIPSMKRTFNLFRITGMPL-----KYLDGEN 175
QY 205 ATWFGLYNISGAGFLHVGLELVN-----HKALDPARMTIQKYFYGGRTYDSIA 256
DB 176 TP--QLYNN-----HFEKAVCSDPYVWSVANGTVPDDVYDVGKQLQAF--GYEKELA 227
QY 257 QLEAOFEGALVNVVLIHDNGTGSMSLKSVPBPAPLQFYRQGRFVSQSRVASSLM 316
DB 228 E--DDKGFDELMVDMMTT--REYLKRGGRGEAPKYDF-----FALQMNETON--- 273
QY 317 TFSFGIAGFSGPRIFDVRFQGERLYEISLQ-----EALAIYGSNS 357
DB 274 -----TGTNLFDQAF--SESVIDSPEDNPTKPEWCIEGSGTS 309

RESULT 12

US-09-912-020-364
Sequence 364, Application US/09912020
Patent No. US20020045592A1
GENERAL INFORMATION:
APPLICANT: Zyskind, Judith
APPLICANT: Ohlsen, Karl L.
APPLICANT: Trawick, John
APPLICANT: Forsyth, R. Allyn
APPLICANT: Froelich, Jamie M.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
FILE REFERENCE: ELITRA.001DVI
CURRENT APPLICATION NUMBER: US/09/912,020
CURRENT FILING DATE: 2001-07-23
PRIOR APPLICATION NUMBER: 09/492,709
PRIOR FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: 60/117,405
PRIOR FILING DATE: 1999-01-27
NUMBER OF SEQ ID NOS: 485
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 364
LENGTH: 878
TYPE: PRF
ORGANISM: E. Coli
US-09-912-020-364

Query Match 2.3%; Score 93.5; DB 10; Length 878;
Best Local Similarity 20.2%; Pred. No. 28;
Matches 108; Conservative 61; Mismatches 190; Indels 175; Gaps 27;

QY 157 YLDIDQIFNRELPOA-----SGLHCCFYKRGHNLVTMT--- 193
DB 148 HLDVGOQRLNLTIPQAFMSNRARGYIPPELMDPCINAGLNN---YNSGNSVQNRIGN 203
QY 194 --TAPRLQSG-----DRATWFGLYNN-----ISAGFFLHVGLELVNKKALDP- 237

DB 204 SHYATLNQSGNLGAMRLRDNTW---SYNSDRSSGSKNKKQHINTFL---ERDIIPL 257
QY 238 -ARMTIQKYFYGGRYDSI-----AOLEAOFEGALVNVVLIIDPNCG--- 278
DB 258 RSRRLTGDDYTGDIIDFGINFRGAQLASDNN-----MLPDSKGFAPVHIGIARTAQ 310
QY 279 -----GWSLSKSPVPPGAPPLQFYRQGRFSYQ-----GS-----R 310
DB 311 VTIKONGDIYNSTVPPGFTINDIYAAGNSGDLQVTTIEADGSIQIFTVPYSSVPLQOR 370
QY 311 VASLSMTFSFGIAGFSGPRIFDVRFQGERLYEISLQEALAIYGSNSPAAMTTRVYDGGF 370
DB 371 EGHTRISITAGEYRSGNAQDEKTRFQSTLLN--GLPAMWTYICGQ---LADRRARWF 425
QY 371 GNGKTTPLTRGVDCPYLATYVDWHEFLSQAPKTRDAFCVPEQNOGLPRRHNSDLYS 430
DB 426 GICKNNAGLG-----ALSYDM-----TOANSTLPDD-----SQHDGQSVR---FLYN 464
QY 431 HYFGGLAEVLVVRSMSTLNDYVDVFEH-----PSGALIRFPYATGYISSAPL 481
DB 465 KSLNESGTNIOLVGYRYSIGYFNEDITYSRNGINIEQSGVIOVAKRFTDYVLA--- 522
QY 482 FGATGKYGNQVSEHTLG---TVHTSAH-----FKYDLVAGLEMMVMAEDMY 526
DB 523 YNKRGLQLQLTVNQ--LGRSTLYLSGSHQTYMCTSNVDEQFALNTA-----FED-- 572
QY 527 FVPMAVPSPEHOLQRLQVTRKLEMEBOAFLVGSATRIYLYLASNHNKQH 580
DB 573 -----INMTLYSYL-----TKNAMQGRDOMALNVNIPSHVLRSDSKSOWMH 616

RESULT 13

US-09-925-301-844
Sequence 844, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA106
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 844
LENGTH: 601
TYPE: PRF
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (36)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (64)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (103)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (106)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (152)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (358)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (383)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-301-844

Query Match 2.3%; Score 93; DB 10; Length 601;
Best Local Similarity 22.3%; Pred. No. 18;
Matches 78; Conservative 32; Mismatches 112; Indels 128; Gaps 19;

QY 12 DGEPSQLPHC-----PSVPSA-QPWTIP-----GQSOLFADLSREE 48
DB 218 DTGTEALTPHIMNRLHATSRKSYRPSMEPMELSPFEDVACTEMQSOSGVDLSGDS 277
QY 49 LTNVMTLQRLGP--GLVDAQAQP-----SDNCVFSVELQPPKAAALHLDRGSPPA 102
DB 278 QVS-SCPCSORSRSPDGLKAAEPPRPGSGSPLNAPCEGPP-----GSEPPR 326
QY 103 REALAIFFGROPQPNVSELVAGPLPHPSYMRDVTVER--HGGPL-PYHR--RPVLE--- 154
DB 327 RPPAPRHDGDRKELPREQPLPPGP---GTERSOXTRDGTPEPGFIRPSHREPPVQFGTX 383
QY 155 QEYLDI-----DOMIFNRE-----LPOASGLHHCFCYKRGRLVTM 192
DB 384 DKSDLLVYVGDLSLKAKEKELTASVTEALPYSRDWELLPSAASA-----EPQSKNIDSG 437
QY 193 TTPAPRGLOSDRATWELLYNISAGFFLHVGLLELVNKHAKALDPAKWTIOKVYQGRY 252
DB 438 HCVEPSSSGQR-----LYPEVFF 456
QY 253 DSLAQLAOFEAGLVNVLIPDNG--TGGSMSL---KSP--VPPGPAP 294
DB 457 GSAPSSSQISGAMDSQLPHNSGFGFPGLPHYSQPLYPGPAP 506

RESULT 14

US-10-099-895-33
Sequence 33, Application US/10099895
Patent No. US20020177166A1
GENERAL INFORMATION:
APPLICANT: BERNDT, Michael C
APPLICANT: STOMSKI, Frank C
APPLICANT: LOPEZ, Angel F
APPLICANT: GUTHRIE, Mark A
TITLE OF INVENTION: A BINDING MOTIF OF A RECEPTOR
FILE REFERENCE: 3991/0K379
CURRENT APPLICATION NUMBER: US/10/099,895
CURRENT FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: PCT/AU00/01118
NUMBER OF SEQ ID NOS: 47
SOFTWARE: PatentIn version 3.1
SEQ ID NO 33
LENGTH: 635
TYPE: PRT
ORGANISM: Homo sapiens
US-10-099-895-33

Query Match 2.3%; Score 92.5; DB 9; Length 635;
Best Local Similarity 21.9%; Pred. No. 21;
Matches 70; Conservative 32; Mismatches 107; Indels 111; Gaps 15;

QY 19 LPHCPSPSAQPMWTHRGOSOL-----FADLSREELTANVMTLQRLGPGVDAQA 70
DB 132 LPAPPSIT-KAMGSGOGELOISHEEPAPETISDLRLYELRGPPDPKNSGTGPYI---QL 187
QY 71 RPSDNCVFSVELQPPKAAALHLDRGSP-----PPAREALAIYFQGQPPQPNV 119
DB 188 IARETC--CPALGRPHSASALDOSPCQPTMPQDQPKQSPSEASALTAEGG----- 239
QY 120 SELVAGPLPHPSYMRDVTVERH---GGLPYHRRPV---LFOEYLDIDOMIFNREL--- 169
DB 240 SCILSIGLOPNSYWLQLRSEPDGSLGSGWSMSLPVTYDLPGBAVALGLQCFILDKNV 299
QY 170 -----PQASGLLH-----C-----CYKRGGR 187
DB 300 TCOMOQDDHASGSGFFYHSRACCPDRDRIWENCEEERTNPGLOTPQPSRCHFKNRND 359

QY 188 N-----LYTMTAPRGLOSDRATWELLYNISAGAFPLHV-----GLELVNKA 224
DB 360 SIHLIIVETTPA-----GTVHSYLSGPFMIHQAVRLPPNLMHREISSCHLE 407
QY 235 LD---PARMTIOKVYQGRY 251
DB 408 LEMOHPSSMAAQETCYOLRY 427

RESULT 15

US-09-712-363-247
Sequence 247, Application US/09712363
Patent No. US20020164588A1
GENERAL INFORMATION:
APPLICANT: Eisenberg, David
APPLICANT: Rotstein, Sergio H.
APPLICANT: Marcotte, Edward M.
TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
FILE REFERENCE: 07419-032001
CURRENT APPLICATION NUMBER: US/09/712,363
CURRENT FILING DATE: 2000-11-13
PRIOR APPLICATION NUMBER: PCT/US00/02246
PRIOR FILING DATE: 2000-01-28
PRIOR APPLICATION NUMBER: 60/179,531
PRIOR FILING DATE: 2000-02-01
PRIOR APPLICATION NUMBER: 60/117,844
PRIOR FILING DATE: 1999-01-29
PRIOR APPLICATION NUMBER: 60/118,206
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: 60/126,593
PRIOR FILING DATE: 1999-03-26
PRIOR APPLICATION NUMBER: 60/134,093
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/134,092
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/165,124
PRIOR FILING DATE: 1999-11-12
PRIOR APPLICATION NUMBER: 60/165,086
PRIOR FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 292
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 247
LENGTH: 904
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-712-363-247

Query Match 2.3%; Score 92; DB 9; Length 904;
Best Local Similarity 19.7%; Pred. No. 40;
Matches 132; Conservative 92; Mismatches 255; Indels 192; Gaps 37;

QY 155 QEYLDIDOMIFNRELPOAS-----GLHHCFCY-KHGRNLTMTTAPRGLO 200
DB 8 KRFIDHVKAGHNEVPSASVILDPNLLFVNAGMVGQVFPFLQGRPPPYTATNSIDKICR 67
QY 201 SGD-----RATWGLYINISGAGFFLH---HYGLELVNKA-----LDPAR-WTI 242
DB 68 TPIDIEYGITRHTNTEFOMANGNSFGDYFRGAIEELAMALLTNSLAAGGYGLDPERIWT- 126
QY 243 QKVEYQG-----RYDSLAQLEAOFEAGLVNVLIPDNGTGS--WSLKSVPVPPGPAPLOF 297
DB 127 -TYFPDDEAVRLMOEYAGLPALF-----RIORGMADNTWSKIPQCGPSSSEI-Y 175
QY 298 YPOGPRFSVQGSRYAS-----SLMTFSFGIAGSGRIPDFVROGERLYEISLAQALAI 352
DB 176 YDRGPEFGAGPIVSDRYLEVNVLYF-----MQRNR--GEGTKEDYQI 219
QY 353 YGNSPAAMTTRYVDGFGMGK-----YTPPLTRGVDCPIYATYV----- 392
DB 220 LG-----PIPRKNIDTGMGVERIALVLYDVHNYETDLR---PVITFARVARAYADV 270

```
OY 393 -----DHFFLESQPKT-----IDAPGVFEQNOGLPLRR-----HHSDLXSHYEG 434
      | : : : : | : | : | : |
Db 271 GNHEDVARYRIADHSRTAAIILIGDVSPGNDGRGYLRLLRRYIRSAKLLIGIDAIVG 330
      | : : : : | : | : | : |
OY 435 GLAEVTL-----VYRSMSTLNDYVMDTFVHPSPGAIEIRPYATGYISSAFLFGAT 485
      | : : : : | : | : | : |
Db 331 DLMATVRNMGPSYPELVADFEERISRIVAEETAFNRITLASSGRUFEE-VASSTKKSAT 389
      | : : : : | : | : | : |
OY 486 GKYNQVSEHTLGTVH-THSAHFKVDLDVAGLENNVMAEDMVEVPM-----AVPWS 535
      | : : : : | : | : | : |
Db 390 VLSSGDAF-----TLHOTYGFPIELTEMAA-ETGLQYDEIGFRELMAEQRRRAKADAAA 443
      | : : : : | : | : | : |
OY 536 PEHQLQRLQVTRKLL-----EMEQQA-----AFLVGSATPRYLXLASNHSNKMGH 580
      | : : : : | : | : | : |
Db 444 RKHAHADLSAYRELVDACATEFTGFDELRSQARILGIFVDGKRPD-----VVAHGVAAGCA 498
      | : : : : | : | : | : |
OY 581 PRGYRIOML-----SPAGEPLPQSSMARGFSWERYOLAVTORKEEPSSSSVFNQND 633
      | : : : : | : | : | : |
Db 499 GEGQREVELDRTPLIYASGGQIADBGTIISGTGSSEARAATDVOKIAKTL-----550
      | : : : : | : | : | : |
OY 634 PWA--PTVDFSDFINNETIAGKDLVAV--VTAGF--LHIPHA--EDI--PNTVTVG--N 680
      | : : : : | : | : | : |
Db 551 -WVHRVNVESGEFEVGDVIAAVDPGWRGATQGHSGTHMVHAALROYLGPNAVQAGSLN 609
      | : : : : | : | : | : |
OY 681 GVGFFLRPNF 691
      | : : : : | : | : | : |
Db 610 RPYGLRPFDFNW 620
```

Search completed: May 20, 2003, 12:09:46
Job time : 29 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: May 20, 2003, 11:56:29 ; Search time 20 Seconds
(without alignments)
3576.200 Million cell updates/sec

Title: US-10-081-408-2_COPY_20_763
Perfect score: 4010
Sequence: 1 LVCVLLVGRGSGDGPSPQLP.....QAAACAPDLPAFSGHGFSSHN 744

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4010	100.0	763	2 JC5234	amine oxidase (cop
2	3331	83.1	762	2 A54411	amine oxidase (cop
3	1398.5	34.9	746	2 S34656	amine oxidase (cop
4	1394	34.8	751	2 A54053	amine oxidase (cop
5	413	10.3	755	2 B41836	amine oxidase (fla
6	397.5	9.9	757	2 E64889	amine oxidase (cop
7	377	9.4	684	2 A56102	amine oxidase (cop
8	368.5	9.2	660	2 G90330	amine oxidase (cop
9	363	9.1	648	2 A48646	amine oxidase (cop
10	357.5	8.9	756	2 T48139	copper amine oxida
11	353.5	8.8	650	2 G71412	probable amine oxi
12	340	8.5	638	2 JC2139	phenylethylamine o
13	325.5	8.1	674	2 JC7251	amine oxidase (cop
14	324	8.1	687	2 T47403	amine oxidase-like
15	320	8.0	674	2 A44239	amine oxidase (cop
16	302.5	7.5	759	2 E84854	probable copper am
17	300	7.4	660	2 AH2234	copper amine oxida
18	295	7.4	692	2 S04963	amine oxidase (cop
19	247	6.2	794	2 T39171	probable peroxidom
20	246.5	6.1	300	2 T48138	copper amine oxida
21	229	5.7	587	2 S21139	amine oxidase (cop
22	214.5	5.3	670	2 S71320	amine oxidase (cop
23	126	3.1	862	2 T46289	hypothetical prote
24	125.5	3.1	460	2 T48137	copper amine oxida
25	125	3.1	757	2 C84120	subtilisin-type pr
26	111	2.8	1332	2 D82685	phosphoribosylform
27	107	2.7	435	2 A42672	cholesterol kinase (EC
28	106.5	2.7	896	2 F96523	hypothetical prote
29	106	2.6	3938	2 T42761	basoon protein -

30	103	2.6	2224	1 KFHU5	coagulation factor
31	102.5	2.6	591	2 C48364	cytochrome-c oxida
32	101.5	2.5	546	2 B75573	conserved hypothet
33	101.5	2.5	848	2 C65083	hypothetical prote
34	101.5	2.5	626	2 T35669	hypothetical ATP/G
35	100.5	2.5	654	2 S76870	hypothetical prote
36	100.5	2.5	1257	1 A41060	neural cell adhesi
37	100.5	2.5	1544	2 E59431	phosphoinositide-b
38	100	2.5	919	2 S42842	T16612.2 protein -
39	99.5	2.5	579	2 E83144	hypothetical prote
40	99.5	2.5	2352	2 C83229	probable non-ribo
41	99	2.5	602	2 T35782	probable secreted
42	99	2.5	915	2 T03589	probable aspartate
43	99	2.5	2055	2 T00093	hypothetical prote
44	98.5	2.5	881	2 S25445	nitrate reductase
45	97.5	2.4	538	2 S76085	hypothetical prote

ALIGNMENTS

RESULT 1					
JC5234	amine oxidase (copper-containing) (EC 1.4.3.6) AOC2 Precursor - human				
C:Species: Homo sapiens (man)					
C:Date: 13-Mar-1997 #sequence_revision 13-Mar-1997 #text_change 18-Feb-2000					
C:Accession: JC5234					
R:Zhang, X.; McIntire, W.S.					
Gene 179, 279-286, 1996					
A:Title: Cloning and sequencing of a copper-containing, topa quinone-containing mon					
A:Reference number: JC5234; M0ID:97128319; PMID:8972912					
A:Contents: placenta					
A:Accession: JC5234					
A:Status: preliminary					
A:Molecule type: mRNA					
A:Residues: 1-763 <RHA>					
A:Cross-references: GB:039447; NID:g1399031; PIDN:AC50919.1; PID:g1399032					
C:Comments: This enzyme catalyzes the oxidation of primary amines to the correspondi					
Cu(II) and 1 molecule of covalently-bound topa quinone. It also can scavenge circ					
C:Genetics:					
A:Gene: GDB:AOC2; DAO2					
A:Cross-references: GDB:4562632					
A:Map position: 17q21-17q21					
C:Superfamily: amiloride-binding protein					
C:Keywords: copper; metalloprotein; oxidoreductase; quinoprotein; topaquinone					
F:1-19/Domain: signal sequence #status predicted <SIG>					
F:20-763/Product: amine oxidase (copper-containing) #status predicted <MAT>					
F:444,520,522/Binding site: copper (His) #status predicted					
F:471/Modified site: topaquinone (Tyr) #status predicted					
Query Match					
Best Local Similarity 100.0%; Score 4010; DB 2; Length 763;					
Matches 744; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
OY	1	LVCVLLVGRGSGDGPSPQLPHCPSPVSPSAQPTWTHPGSQSLFADLSREELTVMFRLQRL	60		
DB	20	LVCVLLVGRGSGDGPSPQLPHCPSPVSPSAQPTWTHPGSQSLFADLSREELTVMFRLQRL	79		
OY	61	GPGLVDAQAQRPSPNCVSVELQLPKRAALAHNDRGSPPARFALIVFGROPOPVS	120		
DB	80	GPGLVDAQAQRPSPNCVSVELQLPKRAALAHNDRGSPPARFALIVFGROPOPVS	139		
OY	121	ELVVGPPLPHSPYRDVVERHGGPLPYHRRPVLFOEYLDIDOMTFNRELPOASGLHNHC	180		
DB	140	ELVVGPPLPHSPYRDVVERHGGPLPYHRRPVLFOEYLDIDOMTFNRELPOASGLHNHC	199		
OY	181	FYKRGRLVTMTTAPRGLSGDRATWEGLYNNSGAGFLLHNGLELVNHNKALDPRRW	240		
DB	200	FYKRGRLVTMTTAPRGLSGDRATWEGLYNNSGAGFLLHNGLELVNHNKALDPRRW	259		
OY	241	TIOKVPFGYGRYDSLAOLEAGFLVNVLIIPNGTGSSTLSPVPPGAPPLQGFYQ	300		
DB	260	TIOKVPFGYGRYDSLAOLEAGFLVNVLIIPNGTGSSTLSPVPPGAPPLQGFYQ	319		

Oy 301 GREFSVQSGSVASSLMFFSGFLGAFSGPRIFDVROGSRLYVEISLQELALATYGGSPAA 360
 Db 320 GPRFVSQSGSVASSLMFFSGFLGAFSGPRIFDVROGSRLYVEISLQELALATYGGSPAA 379
 Oy 361 MTRTRYVDGFGGMAKYYTTPPLTRGRVDCPYLATYVDHMFLLLESQAPKTRDAFCVFEQNGLP 420
 Db 380 MTRTRYVDGFGGMAKYYTTPPLTRGRVDCPYLATYVDHMFLLLESQAPKTRDAFCVFEQNGLP 439
 Oy 421 LRRHNSDLYSHYFGLAEYLVLRVNSKSTLLNYDVMDFVHPSGAIIEIRFYATGYISSAF 480
 Db 440 LRRHNSDLYSHYFGLAEYLVLRVNSKSTLLNYDVMDFVHPSGAIIEIRFYATGYISSAF 499
 Oy 481 LFGATGKYGVQVSEHNTLGTYHTHSAHFRKVDYDVAGLEMMVNAEDVAFVMAVPMSPSEHOL 540
 Db 500 LFGATGKYGVQVSEHNTLGTYHTHSAHFRKVDYDVAGLEMMVNAEDVAFVMAVPMSPSEHOL 559
 Oy 541 ORLQVTRKLLMEEQOAFVLGSAAPRYLYLASNHSNKKGHPRGYRIQMLSFAGEPPLPONS 600
 Db 560 ORLQVTRKLLMEEQOAFVLGSAAPRYLYLASNHSNKKGHPRGYRIQMLSFAGEPPLPONS 619
 Oy 601 SNARGSWERYQLAVTORKEEPPSSSYFNONDPAAPVDFSDFINNETIAGDKLYAWYT 660
 Db 620 SNARGSWERYQLAVTORKEEPPSSSYFNONDPAAPVDFSDFINNETIAGDKLYAWYT 679
 Oy 661 AGFLTHPHADIDPNTVTVGVNGVGFLLRPYNFDEDEPFSFSAISYFRGQDADAGACEVNP 720
 Db 680 AGFLTHPHADIDPNTVTVGVNGVGFLLRPYNFDEDEPFSFSAISYFRGQDADAGACEVNP 739
 Oy 721 ACPLPAAACAPDLPAPFSHGSGFSHN 744
 Db 740 ACPLPAAACAPDLPAPFSHGSGFSHN 763
 RESULT 2
 A:4411
 amine oxidase (copper-containing) (EC 1.4.3.6), serum, precursor - bovine.
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 09-Sep-1994 #sequence, revision 06-Feb-1995 #text_change 18-Feb-2000
 C:Accession: A54411; B38081; A48242; S65408
 R:Mu, D.: James, S.M.; Smith, A.J.; Brown, D.E.; Dooley, D.M.; Klimman, J.P.
 J: Biol. Chem. 267, 7979-7982, 1992
 A:Title: Tyrosine codon corresponds to topa quinone at the active site of copper amine oxidase.
 A:Reference number: A38081; MUID:9235001; PMID:1569055
 A:Accession: A54411
 A:Molecule type: mRNA
 A:Residues: 1-762 <MUA>
 A:Cross-References: GB:S65983; NID:9546215; PIDN:AAB30397.1; PID:9546216
 R:Mu, D.: James, S.M.; Smith, A.J.; Brown, D.E.; Dooley, D.M.; Klimman, J.P.
 J: Biol. Chem. 267, 7979-7982, 1992
 A:Title: Tyrosine codon corresponds to topa quinone at the active site of copper amine oxidase.
 A:Reference number: A38081; MUID:9235001; PMID:1569055
 A:Accession: B38081
 A:Molecule type: protein
 A:Residues: 463-465, 'D', 467-473, 'X', 475-485 <MU2>
 R:James, S.M.; Mu, D.; Wemmer, D.; Smith, A.J.; Kaur, S.; Maltby, D.; Burlingame, A.L.;
 A:Title: A new redox cofactor in eukaryotic enzymes: 6-Hydroxydopa at the active site of
 A:Reference number: A48242; MUID:90260648; PMID:2111581
 A:Accession: A48242
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 468-472 <JUA>
 R:de Blaas, D.; Agostinelli, E.; de Matteis, G.; Mondovl, B.; Morpurgo, L.
 Eur. J. Biochem. 237, 93-99, 1996
 A:Title: Half-of-the-sites reactivity of bovine serum amine oxidase. Reactivity and chem
 A:Reference number: S65408; MUID:96203913; PMID:8620899
 A:Accession: S65408
 A:Molecule type: protein
 A:Residues: 463-469, 'X', 471-487 <DEB>
 C:Superfamily: aniloride-binding protein
 C:Keywords: copper; glycoprotein; oxidoreductase; cupinoprotein; topaquinone
 F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-762/Product: amine oxidase (copper-containing) #status predicted <MAT>
F:136,221,665/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:470/Modified site: topaquinone (Tyr) #status experimental

Query Match	83.1%	Score 3331	DB 2	Length 762
Best Local Similarity	81.4%	Pred. No. 2.2e-257		
Matches 609; Conservative	59;	Mismatches 70;	Indels 10;	Gaps 2;

QY	4	VLLVR--GGGGG-----PSQLPHCPSPVSPAOQWTHPGSQULPADI.SREELTAVM	53
Db	12	LLVMKREGGCGSEEGVCKOCHPSLPPRCPSKSPSDQWTHPDQSOLPADI.SREELTAVM	71
QY	54	RELTORLPGVLDAQAARSPDNCVFSVELQLPCKAALAHLDGSPPREARELAIVFGR	113
Db	72	SFLVQLQLPDLVDAQAARSPDNCVFSVELQLPCKAALAHLDGSPPREARELAIVFPG	131
QY	114	QOPVNSSELVYGRPLPHPSYMDYVERHGGRPLPHRRPVYLEOETIDIDQMFNRLPAS	173
Db	132	QOPVNTVELVYGRPLPQPSYMRDYVERHGGRPLPYRRPVLLREYDIDIDQMFNRLPQA	191
QY	174	GLLHHCCGYKHHGRLLYMTTAPRGLOSGDRATWYGLYNNISGAGFPLHHGLVYHNK	233
Db	192	GLVHHCCSYKGGQLLLTNNNSAPRGVQSGDKSTWFGITYNNTKGGPLPHPGLELDYHK	251
QY	234	ALDPARWTIQKYFYOGRRYDSLAEQAEAGLVNVLIPDNGTGSWSLSPVPPGPAP	293
Db	252	ALDPADMTYQKYFQGRYYENLAQLEQFEGAGVNVVVIIPDDGTGFWSLKSQVPPGPTP	311
QY	294	PLQFTPGQPRPESVQSSRAVSSLMTFSFGLGAFSGRITIDYRQGRLLYETSLDALIY	353
Db	312	PLQFHPQGRFRESVQSNRAVSSLMTFSFGLGAFSGPRFVDYRQGRLLAYEISLOBAGAY	371
QY	354	GGNSPAAMTRTVVDGFGFGKTYTPLTRGVDCPYATATVVDHFILESOAPKIRPACVF	413
Db	372	GGNTPAAMTRTVMDSGFNGFYATPLIRGVDCPYATATYMDHFIYVESOTPKTLHAPCFV	431
QY	414	EONOGPLPRRHSDLYSHYFGGLAETVLVRSMTLLNYDVMYVTFPHSGAIEIRFAYT	473
Db	432	EONKGPLPRRHSDLSHYFGGVAQTVLYFRSVMLNADYVMAYVFPNGAIEVKLHT	491
QY	474	GYSSAFLEFGATGKGNVSEHTLGTVHTSHAFVYDLDVAGLEMMVVAEDMYFPAAPV	533
Db	492	GYSSAFLEFGAARRRGNQGEHTLGPVHTSHAFYVDLDVGGLEMMVVAEDMAFYPTAIP	551
QY	534	WSPERQIOLVOTRKLLEMEQAAFLVGSATPRYLYLASNHSNKKMGHGRGVYIOMLSFAG	593
Db	552	WSPERQIOLVOTRQLEHEQAAFLPGASPRYLYLASKOSNKKMGHPRGVYIQVSPAG	611
QY	594	BELPONSNAARGFSEMYOLATVORKEEPPSSSVFNQNDPAPVAVDESDFINNETIAGK	653
Db	612	GMPONSPIERAF.SMGRIOLATVORKEETPPSSSVFNQNDPPTPYVDSDFINNETIAGK	671
QY	654	DLVAVYTAGFLHI.PHAEDIPNTVTVNGVNGVFFLARYNFFEDDEPSEY.SADSIYFRGDODAG	713
Db	672	DLVAVYTAGFLHI.PHAEDIPNTVTVNGVNGVFFLARYNFFEDQPSMDSADSIYFRGDODAG	731
QY	714	ACEVNPPLACTPAAACAPDLPAFASHGF	741
Db	732	ACEVNPPLACTPAAACAPDLPAFASHGF	759

result 3
 S34656
 amine oxidase (copper-containing) (EC 1.4.3.6), kidney, precursor - rat
 N:Alternate names: amiloride-binding protein, long form
 N:Contains: amiloride-binding protein, short form
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 18-Feb-2000
 C:Accession: S36847; S36848; IS1904; S34656; S34657
 R:Ringveggli, E.; Renard, S.; Volley, N.; Waldmann, R.; Chassande, O.; Lardunski, M.
 Eur. J. Biochem. 216, 679-681, 1993
 A:Title: Molecular cloning and functional expression of different molecular forms of
 A:Reference number: S36847; MUID:93387321; PMID:83375402

Query Match	34.98;	Score 1398.5;	DB 2;	Length 746;
Best Local Similarity	40.38;	Pred. No. 3.3e-103;		
Matches 294;	Conservative 127;	Mismatches 266;	Indels 43;	Gaps 13;

```

Db      676  DVPNTAATPGNSVGLTLRPFNFEEPPEDPSLASRDTIV-V-PPDQKGLNRVO--RWIPEDRRCL 732
Oy      731  PDLPAPSHGC 740
Db      733  VS-PPEFSG 741

RESULT 4
A:Accession: A54053
A:Name: amine oxidase (copper-containing) (EC 1.4.3.6) ABP1 precursor - human
A:Alternate names: amiloride-binding protein; diamine oxidase; histaminase
C:Species: Homo sapiens (man)
C:Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 18-Feb-2000
C:Accession: A54053; A54410; I38708; A38276; S42495
R:Chassande, O.; Renard, S.; Barbry, P.; Lazdunski, M.
J. Biol. Chem. 269, 14484-14489, 1994
A:Title: The human gene for diamine oxidase, an amiloride binding protein. Molecu
A:Reference number: A54053; MUID:94237856; PMID:8182053
A:Accession: A54053
A:Molecule type: DNA
A:Residues: 1-751 <CHAB>
A:Cross-references: EMBL:X76212; NID:9463242; PIDN:CA55046.1; PID:9463243
R:Novotny, W.F.; Chassande, O.; Baker, M.; Lazdunski, M.; Barbry, P.
U. Biol. Chem. 269, 9921-9925, 1994
A:Title: Diamine oxidase is the amiloride-binding protein and is inhibited by amil
A:Reference number: A54410; MUID:94193685; PMID:8144586
A:Accession: A54410
A:Molecule type: protein
A:Residues: 20-27 'A', 29-39 <NOV>
A:Cross-references: PIDN:AB30395.1; PID:9546213
A:Experimental source: Placenta
A:Note: sequence extracted from NCBI backbone (NCBIF:146046)
R:Zhang, X.; Kim, J.; McIntire, W.S.
Biochem. Genet. 33, 261-268, 1995
A:Title: cDNA sequences of variant forms of human placenta diamine oxidase.
A:Reference number: I38708; MUID:96113540; PMID:8595053
A:Accession: I38708
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-331, 'F', 333-751 <RES>
A:Cross-references: EMBL:U11862; NID:9533335; PIDN:AAC50270.1; PID:9533336
R:Barbry, P.; Champe, M.; Chassande, O.; Munetitsu, S.; Champigny, G.; Lingueglia
Proc. Natl. Acad. Sci. U.S.A. 87, 7347-7351, 1990
A:Title: Human kidney amiloride-binding protein: cDNA structure and functional exp
A:Reference number: A38276; MUID:91017502; PMID:2217167
A:Accession: A38276
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-265, 'RCGARGMTAQRSPSSPPQAPRDFPPHPCEPPLGAPAPRPSLAAGGCRALRLELPLV'
A:Cross-references: GB:M56602; GB:M6335; NID:9387655; PIDN:AAA58358.1; PID:g17796
A:Note: this sequence has been corrected in reference A54053
C:Genetics:
A:Gene: GDB:ABP1
A:Cross-references: GDB:127105; OMIM:104610
A:Map position: 7q31-7q32
A:Introns: 524/1, 619/2, 663/3
C:Function:
A:Description: oxidation of putrescine and histamine
C:Superfamily: amiloride-binding protein
C:Keywords: homodimer; oxidoreductase; quinoprotein; topaguinone
F.1-19/Domain: signal sequence status predicted <Sto>
F.161/Modified site: topaguinone (Tyr) #status predicted

Query Match      34.8%; Score 1394; DB 2; Length 751;
Best Local Similarity 39.8%; Pred. No. 7.6e-103;
Matches 293; Conservative 124; Mismatches 281; Indels 38; Gaps 11;

Oy      31  PWHPGQSOLFADLSKEELTAVWRFLTORLGPGLVDAQAARPSDNCVFSEVLEQLPKAA 90
Db      23  PGLPLPKAGVFSDSLNOELKAVHSFLMSKKEELRQPSSTTTMAKNVFLLEMLLPKKYHV 82
Oy      91  LAHLIDGSPPARREALALVYFPGQOPQPNNSLVVGLPLPHSYKRDVTVERHGSRPLRYNHR 150

```

[illegible]

Oy 56 LTORAGPELVDAQAARSDNCVFSVELQPKAAALHLDGSGPPA-REALATVFFGRQ 114
 Db 140 -----IYAAAEFPQPNTRFTEISLHEPDKAAVMAFALGITEVDAPRTADVYMLDGK- 190
 Oy 115 POPNVELVGPRLPHPSFMRDVTVE-RHGGLPYRRPRLVEQELIDIDOMIFNELPQAS 173
 Db 191 ---HYIEAVV-DLQNKKILTSWPIKGANG-----MVLDDPVSQNIH----- 229
 Oy 174 GLHHCCYTKRGRNLYMTTAPRGLOGSDPATFGLYTNISGAGFLHGHVLELVNHNK 233
 Db 230 -----NNSSEFA-----EVLKHKG 243
 Oy 234 ALDPARWITOKV---FYQGRYDLSLAOLEAQFEAGLVNVLIPNGTGSWS----- 282
 Db 244 ITDPEKVVTTFLTVGFEDGK--DGLQO-----DARLVVSYLDTGCGNTAHPHLEMLVA 296
 Oy 283 ---LKS-----PVDPGPAP-----PLQF-YQGPFRFSVQSGRVAAS 314
 Db 297 VVDLEAKKIILEGPVIVPMEPRPVDGRDRNAPAVKFLITEPECKNNTITGDTIHMQ 356
 Oy 315 LMTFSFGAGRSGRPIFPVRFG---ERLYVEISLQALAYGNSNAAMTTRYVNDG- 369
 Db 357 NMDFLRNLSRVGPLSTVTVTNDGKTRQVMEESLGMIVPYGDPDVGWYFKAYLDSGD 416
 Oy 370 FGMGKTYTPLTRGYDCPLATVYVDMHFLLESQAERTIDACVFEONGLRLRRHSDLY 429
 Db 417 YGMGLTSPPIRYRGADAPBNAVLDETTADYTGKPTTIGCAVAIE-----KAGPEY 468
 Oy 430 SHYEGG----LAETLVLRSMSTLLNTDYVDVTFHPSGAIEIRFATYGYISSAFLEGA 484
 Db 469 KHELGKPNVSTERRELVAWISTGVNGDYDIFDWVHNGTIGIDAGCGIEAVKGLAK 528
 Oy 485 T-----GRYGNQVSEHNLGTVHTHSAPFKVDLDVAGLENNYMAEDMVFVPAPVMS 535
 Db 539 TMDHPSAKEDTRYGTLIDHNITVGTHQHINFRDLVDVGGNNLVLAADPEVKRPT--A 585
 Oy 536 PEHQRLQVTRKLLLEMEEOAAFLVGSATPRILYLASNHS--NKMGHPRGRIOMLSPAG 593
 Db 586 GGRPRSTQVQVQNTTDSOKAKAAOKFDPGTLIR---LLSTSKENRGNQNVSY--QIIIPAG 640
 Oy 594 --EPILQNSMA---RGFSWERYOLATVQRKEE-----PSSSVFNQNDPMAPT 638
 Db 641 GTNHPAATGAKFAPDEWIMYHRLSFMOKQLMVRHYHPTERYPEGKYPNRBAHDPTGLQGYAK- 699
 Oy 639 VDFSPFINNETIAGDLVAMVTAGFLHLPNHEIDIPNTVYVNGVGFRLRPNTFDEDS 697
 Db 700 -----DDESLTNHDVWITTTGTHVARAEWIMPPT--EMALLALKPNWFDEPT 749

RESULT 6
 EG4889
 E:amine oxidase (copper-containing) (EC 1.4.3.6) tylna precursor - Escherichia coli (str
 N:Alternate names: monomamine oxidase; tyraminase; tyramine oxidase
 C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence,revision 17-Sep-1997 #text_change 01-Mar-2002
 C:Accession: EG4889; I40923; S65442
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
 A.: Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:9742617; PMID:9278503
 A:Accession: EG4889
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-757 <BLAT>
 A:Cross-references: GB:AE000235; GB:U00096; NID:g1787643; PIDN:AACT74468.1; PID:g17876
 A:Experimental source: strain K-12, substrain M61655
 R:Azakami, H.; Yamashita, M.; Roh, J.
 J. Perment. Bioeng. 77, 315-319, 1994
 A:Title: Nucleotide sequence of the gene for monomamine oxidase (maoA) from Escherichia
 A:Reference number: I40923
 A:Accession: I40923

A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1,247, 'E', 249-257, 'V', 260-276, 'I', 277-287, 289, 'I', 291-455, 'P', 457-658, 'D', 6
 A:Cross-references: GB:23670; NID:9474301; PIDN:BA04900.1; PID:9809499
 A:Experimental source: strain K-12; substrain W3110
 R:Steinbach, V.; Benen, J.A.E.; Bader, R.; Postma, P.W.; de Vries, S.; Duine, J.A.
 Eur. J. Biochem. 237, 584-591, 1996
 A:Title: Cloning of the *macA* gene that encodes aromatic amine oxidase of *Escherichia coli*
 A:Reference number: S65442; MUID:96235221; PMID:8647101
 A:Accession: S65442
 A:Molecule type: protein
 A:Residues: 41-48 <STE>
 A:Experimental source: strain W3350
 C:Genetics:
 A:Gene: *lynA*
 C:Complex: homodimer
 C:Function:
 A:Description: catalyzes oxidation of phenylethylamine and water to phenylacetaldehyde,
 A:Pathway: amino acid metabolism; phenylethylamine catabolism
 A:Note: 2,4,5-trihydroxyphenylalanine quinone (topaquinone) cofactor; copper cofactor; h
 C:Superfamily: amine oxidase (copper-containing)
 C:Keywords: copper; oxidoreductase; quinoprotein; topaquinone
 F:1-33/Domain: signal sequence #status predicted <SIG>
 F:34-757/Product: amine oxidase (copper-containing) #status predicted <MAT>

Query Match 9.9%; Score 397.5; DB 2; Length 757;
 Best Local Similarity 27.5%; Pred. No. 2,5e-23;
 Matches 142; Conservative 74; Mismatches 211; Indels 89; Gaps 20;

QY 246 FVQGRYDLSLQLEAFEGVNVVLIIPDNGTSGS-----LKS 285
 DB 259 YEDGK--DELKQ-----DRLKLVISYLDVGDGNTYANHTENLVANVDEQKIVIEEG 311
 QY 286 PVPKPG-----AP---PLOGY--POGPRFSVQSRVASSLMTFSFGIGAFSGPR 329
 DB 312 PVVPRVMTAPRFGGRVAPVAKPMQIIEPEGKNTITGDMIMRMDFLHNSNRVGP 371
 QY 330 IFDVRFQG---ERLYEISLQALAIYGGNSPAAMTTRVVDG--FGMKYTTPLRGVD 384
 DB 372 ISTVYNDNGTKRKVMYEGSLGMIVPYGDPDGMWTFKALDSDGDMGTLSPIAGKD 431
 QY 385 CPVLATVYVDMHLELSQAPRTIRDAFCVFQNGGLPLRHNSDLYSHYFG-----LAET 439
 DB 432 ABSNAVLNHTTADYTGVPMEIRALVFE-----KAGREYKQENQGRNVSERR 483
 QY 440 VLVRKSTLNDYVDYFHFSGAIEIRFATGY-----ISSAFLEGATG---YGN 490
 DB 484 ELYVRKISTGVNDYIFEDMIFHENGSTIGDAGATGIEAVGVAKATWHDFAKDDTRGYT 543
 QY 491 QVSEHTLGTVHTSAHFKVDLDVAGLENNVMAEDMVFVPAVWSPRPHQDRLDVTKL 550
 DB 544 LIHNIITGTHOHLYNFRDLDDGNNNSLVANDPVVAKPNT---AGSPRTSTMVQNYNI 600
 QY 551 EMEQAAFLVGSATPRRYLYLASNHSNKGPRGRIOMLSFAG--EPLPQNSAA----- 603
 DB 601 GNEDDAAQKDFDPTIR--LSPNPKENKGNPVSI--QIIPRAGSTHYAKAQAQAPDEMI 657
 QY 604 -RGSFWRVOLAVYQKKEEPSSSSVF--NONDPWAPTVDSFDTNNTIAGKDLVAVNTA 661
 DB 658 YHRSEFDKDLWYTRYHGRGEFPGKYPNRSTHDTGLQYSK--DNESLDNTDAVVMWTT 715
 QY 662 GFLLHPIHAEDIPNTVTVGNGVGFLLRYNPFDEPS 697
 DB 716 GTTHVAREBWPIMPT--EWVHTLLKPMNFEDEPT 749

RESULT 7
 A56102
 amine oxidase (copper-containing) (EC 1.4.3.6) - *Arthrobacter globiformis*
 C:Species: *Arthrobacter globiformis*
 C:date: 13-Apr-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000
 C:Accession: A56102
 R:Choi, Y.H.; Matsuzaki, R.; Fukui, T.; Shimizu, E.; Yorifuji, T.; Sato, H.; Ozaki, Y.;

J. Biol. Chem. 270, 4712-4720, 1995
 A:Title: Copper/topa quinone-containing histamine oxidase from *Arthrobacter globifo*
 A:Reference number: A56102; MUID:95181469; PMID:7876243
 A:Accession: A56102
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-684 <CHD>
 A:Cross-references: GB:D8508; NID:9994746; PIDN:BA07517.1; PID:9994747
 A:Note: authors translated the codon TIG for residue 239 as phe
 C:Superfamily: amine oxidase (copper-containing)
 C:Keywords: oxidoreductase

Query Match 9.4%; Score 377; DB 2; Length 684;
 Best Local Similarity 24.1%; Pred. No. 9.1e-22;
 Matches 182; Conservative 107; Mismatches 343; Indels 124; Gaps 28;

QY 20 PRCPSVPSAQWMT--HPCQSOLFADLSREELAVVRFLLQRLGSLVDAQAARPSNCV 77
 DB 7 PESTPLVQDPVPVATLVHAAQHLELSAEIHEARRILAE--AGLVG----- 52
 QY 78 FSVIELPRKAAALAHLDNGSP-----PAEALAIYFGRGQ-----PNVSE 121
 DB 53 -----ESTRFAYGLIEPPKTRRGQDVYGAARLVAMIPRAAQSLSLVRLSLAIG 103
 QY 122 LVVGRPLRHSYMRDVTVERHGGPLRYHRRVLEFOEYLDIDOMIFNELPQASGLHNCSF 181
 DB 104 LVVD-----RRELNPEDAG-----QLPVLEEGLIEDL--SEDPQMNAL----- 143
 QY 182 YKHGRNLYMTAP-----RGLOSQRATWGLYVNSGAFGLHNGLELVNHEKA 234
 DB 144 -TAGGLTPAQVRAVRLSAGVFEYGENEGKRL-----LRGSGFQNDHP-----ADHPW 189
 QY 235 LDPARKTIQVYFQGRYDLSLQLEAFEGVNVVLIIPD--NGTGSWSLSKSPVPGAP 293
 DB 190 AHPIDILVAVDVENRRVNL-----IDGPRPV--PVRNGYTDPRALRGELRIDL 240
 QY 294 PLOGYPOGPRFSVQSRVASSLMTFSFGIGAFSGPRIFDVR--FOGER--LYEISLQEA 349
 DB 241 IETMOPRGSFTLEGNHLSAGNDLRYGPRAREGLVHOHNSHKGRRRVYINRASESEM 300
 QY 350 LAIYGGNSPAAMTTRYVDG--FGMKYTTPLTRGVDCRYLATIYDMMHFLLESQAPKTRD 408
 DB 301 VVPGDPSPYRSMQNYFDSGEYLGRDANSRLDCDGLGYTVPVADDFGNPKTEEN 360
 QY 409 AFCVFQNGGLPLRHNSDLXSHYFGSLAEVLYVVRKSTLNDYVDYFHFSGAIEI 468
 DB 361 GICIHEDAGILMK--RTDEWAGSDEVRRNRRLVVSFTTYGNDYGFYTWLYLDGTIEF 418
 QY 469 RFVATGYISSAFLEGATGKGNQVSEHTLGTVHTSAHFKVDLDVAGLENNVMAEDMVF 528
 DB 419 EAKTGTIVFTALPDKRYAYASEIARIGAYHOLHSARIDMMIDGQANVEELDYLRL 478
 QY 529 PMAVWSPREHOLQVYTRKLEMEQAAFLVGSATPRRYLYLASNHS--NKGHNRGRGRIQ 587
 DB 479 PKG--PGNPHG--NAFTQKRTLLARESEAVRDADAKGVWISIPDLSLHNGHVPVGYLY 535
 QY 588 MLSFAGER---LPQNSMARGFSWERYOLAVYQKKEEPSSSSVFNONDPWAPTVDSDF 644
 DB 536 P--EGNPTLAMADDSIASRAAFARHNLWTRNAEEELVYAGDFVNOHPCGAVLP--AYV 591
 QY 645 INNETIAGKDLVAVNTAGFLHIPIHAEDIPNTVTVGNGVGFLLRYNPFDEPSFYSADSI 704
 DB 592 AQDDIDIGODLVVHNSFGLTHFRPREDP--IMPVDTGFLTLKPRGFDENPTLVNPPSA 649
 QY 705 YFRGDQAGACEVNPACLPOAAACAPDLPAFSHG 740
 DB 650 -----AGHCGTG-----SERENHAPGTAVVGHSG 673

RESULT 8
 G90330
 amine oxidase (copper-containing) (*lynA*) [imported] - *Sulfolobus solfataricus*

C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001
C:Accession: G90330
R:Shen, Q.; Stieglitz, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Amey, M.J.; Chan, J.; Jongs, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R. et al., R.A.; Ragan, M.A.; Sengen, C.W.; Van der Oost, J.
Submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: G90330
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1,660 <KUR>
A:Cross-references: GB:AE006641; NID:g13814942; PIDN:AAK41902.1; GSPDB:GN00155
C:Genetics:
A:Gene: tynA
C:Superfamily: amine oxidase (copper-containing)

Query Match 9.24; Score 368.5; DB 2; Length 660;
Best Local Similarity 24.44; Pred. No. 4,1e-21;
Matches 175; Conservative 103; Mismatches 303; Indels 137; Gaps 28;

QY 44 LSRELTAAVRFRTORLPGLVDAQAARPSDNCVSVELQPPKAAALAHLDROGSPPAR 103
DB 17 LNEEIKRSVEVLKROLN---LDAKVVK---FFSVELKEPKKOEYLEWNNKNIKIER 68
QY 104 EALIVFEGROPOPNVSELVGPPLPHPSYMDYVERHSGELPHRRPVLFQELIDIDQ 163
DB 69 ESL-VKYYNPODR-KVYEALV---SLDNNVKEIRSIDAAV--PRITLDEGECECA 118
QY 164 IFNELPOASGLHHCFCYKRGRLVMTTAPRGLOSGDRATWGLYINISGAFGLH 223
DB 119 VRNDRKROEA-----LTRGILINDLN---LWVDCWABP---H 151
QY 224 VGELELVNKAALDPAKTIQKRYOGRYDSLAQLEAFAGLVNVLIPDNGTGSWSL 283
DB 152 VDEELGRRAVIG-YMW-VKKDIEDNGYGRVHGLMPVLDKMEVIRIDHGT----- 203
QY 284 KSPRPAPR-----PRLOF-YPOGRFSVGSRAVSSLTWTSFGLGAFSG 327
DB 204 -SPPLDADATTPKLOKIGDGLKPLEIKROPKSSSIKINQWETISYRWRLRIYTPREG 262
QY 328 PRFEDVRFOGER-----LVYEISLQEAALAYGNSPAAATRYVDG-FGCKYTPPLTR 381
DB 263 LVYIDVYIDENNERMILYRASVVDLWVPGDPSPHNNKMYLDAGDYGIGLNTIVLSH 322
QY 382 G-----VDC-PIYATVDMHFLLESQAPKTRIDAFVFEONOGPLRRHSDLYSHYFG 434
DB 323 GNYDLVNCDFGEVITYHLDVTRVSSNGTPRIKIKACVHEEDFGVLMR--HTDLRS--- 376
QY 435 GLAET-----LVYRSMSTLNYDVMDTVFHPGATIEIRYATGYISSAFL--GATGY 488
DB 377 GKSEVRNRRRLVSPMATLNYDGFYFYODSIEFLVKLGTIINDSISEKDPPTY 436
QY 489 GNOVSEHTICTVHTSHAFKVDLDVAGLENNVMAEDVVFVMAVMSPEHOLQIOTYRK 548
DB 437 GTRVTPENVAPRIHHEFNIRLNINVDGLRNRIT-----EVLNREBPITEK 481
QY 549 -----LLEMEQAFLVGSATPRYLILAS--NHSNKGHPRGYRIOMLSFAGER 595
DB 482 NPVGNAPFAEENLEENADARRHNPOTGRYKIVNIQKNKYLGLPYAYRLVPGHNVLP 541
QY 596 LPONSSMARGFSEMYOLAVTORKEEPPSSSVFNQNDMPAATYDFSD-----FINNETI 650
DB 542 LPDDSYRRRGATINHLWMTTPNEERYASGD-----PYLRADGGLPKYTLKKRSI 594
QY 651 AGDLVAMVTAAGFLIHIAEDIPRTVTVGNGVGFLLRPYNFEFDDPSFYSDSIYFRG 708
DB 595 VDEDLVIMVTLGVENHVRIEDMP--VMPVEMAGFRLLPDPFDPKNPITLYLPROLRING 650

RESULT 9
A48646

amine oxidase (copper-containing) (EC 1.4.3.6) - Arthrobacter sp. (strain P1)
C:Species: Arthrobacter sp.
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 29-Sep-1999
C:Accession: A48646
R:Zhang, X.; Fuller, J.H.; McIntire, W.S.
J. Bacteriol. 175, 5617-5627, 1993
A:Title: Cloning, sequencing, expression, and regulation of the structural gene for t
hylotriph.
A:Reference number: A48646; MUID:93374858; PMID:8366046
A:Accession: A48646
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1,648 <ZHA>
A:Cross-references: GB:L12990; NID:g289152; PIDN:AAA22074.1; PID:g289153
C:Superfamily: amine oxidase (copper-containing)
C:Keywords: oxidoreductase

Query Match 9.14; Score 363; DB 2; Length 648;
Best Local Similarity 24.08; Pred. No. 1,1e-20;
Matches 177; Conservative 96; Mismatches 301; Indels 164; Gaps 31;

QY 36 GOSQLFADLSREELTAARVFTORLPGLVDAQAARPSDNCVSVELQPPKAAALAHLD 95
DB 12 GVSHPDLPSVELARAVAILKE--GPAALSF-----FISVLEKPSKDL----- 57
QY 96 RGSPPAREALAIYFEGROPOPNVSELVGPPLPHPSYMDYVERHSGELPHRRPVLFQ 155
DB 58 RAGVAVAREADAVLV--DRAQARFEAVVD-----LEAGTVDSMKLLAENIQPPFLD 108
QY 156 EYLDIDQIFNELPOASGLHHCFCYKRGRLVMTTAPRGLOSGDRATWGLYINISGAFGLH 212
DB 109 EFACEDAC--RKDE-----VIALAKRGILNLDVCEPWSVGYF 148
QY 213 NISGAFGLHVGLELVNKAALD-PAKTIQK--VFQGRYDSLAQLEAFAGLVNVLIPDNGTGSWSL 269
DB 149 GEDNDRGL--MRLVYFRDADSDSPYAHPIENTVYF-----DLNAG--KV 191
QY 270 VLPDNGTGSWSLSKSPVPAPPLQFY-----POGRFSVGSRAVSS 314
DB 192 VRLDDQA-----IPVSARGNYLPRKYGEARTDLKPLNITQPEGASFYVTGNHTYWA 244
QY 315 LWFPSFGIAGSFRIDVROGE-----LVYEISLQEAALAYGNSPAAATRYVDG-FGCKYTPPLTR 381
DB 245 DMSFRVGTTPREGVLVHOLKFKDGVDRPVYINRASLSMVPVYDGLTAVQAKKNAFDSGE 304
QY 370 FGMRKYTPPLTRGVDCPLATVYDMHFLLESQAPKTRIDAFVFEONOGPLRRHSDLYSHYFG 434
DB 305 YNIGMANSLTLGDCGEIKYFPGHSDVSGNWTIENALCMHEEDSI-----LW 356
QY 430 SH--YFGGLAET-----LVYRSMSTLNYDVMDTVFHPGATIEIRYATGYISSAFL--GATGY 488
DB 357 KHDFRGTETATRSRKLVIISFIATVANYEYAFYWHFLDQSIIEFLVKATGILSTAGOLP 416
QY 483 GATGYGNOVSEHTL-GVHTSHAFKVDLDVAGLENNVMAEDVVFVMAVMSPEHOLQIOTYRK 548
DB 417 GKKPPTQSLNDGLVAPRIHHEFNIRLNINVDGLRNRIT-----EVLNREBPITEK 481
QY 541 -ORLOVTRKLEMEQAFLVGSATPRYLILASNHS--NKGHPRGYRIOMLSFAGEBLPQ 598
DB 469 GTAFMAVDRLLETQKAIKRTNEAKHFRFKIANHNSNLVNEPARYL-----IPT 519
QY 599 NSSARGFSEMYOLAVTORKEEPPSSSVFNQNDMPAATYDFSD-----FSDFINNETI 650
DB 520 NG-----IOLAA--RDDAVYSKRAQFAARNLWTAADYTRERPAAGEYPMOATGAD 567
QY 650 -----TAGDLVAMVTAAGFLIHIAEDIPRTVTVGNGVGFLLRPYNFEFDDPSFYSDSIYFRG 708
DB 568 DGLHITQKRNINVDVLDLVVWTFGHHVVRLEDMP--VMPRONIGMLPEHGFENQNP 625
QY 698 FYSADSIYFRGDDAGAC 715
DB 626 LNLPTSTTGTGTGADTC 643

RESULT 10

T48139

copper amine oxidase-like protein - Arabidopsis thaliana

N:Alternante names: protein T4C9.130

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Sep-2000

C:Accession: T48139

R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Men

submitted to the Protein Sequence Database, June 1999

A:Reference number: 224485

A:Accession: T48139

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-756 <BEV>

A:Cross-references: EMBL:AL080318

A:Experimental source: cultivar Columbia; BAC clone T4C9

C:Genetics:

A:Map position: 4

A:Insertions: 460/3; 498/3; 686/2

A:Note: T4C9.130

C:Superfamily: amine oxidase (copper-containing)

Query Match 8.9%; Score 357.5; DB 2; Length 756;

Best Local Similarity 22.2%; Pred. No. 3.0e-20; Indels 191; Gaps 29;

Matches 163; Conservative 100; Mismatches 280;

75 NCVSVELQLPPKAAALAHLDGSPPPAREALAI-----VFGRQOPNVSE 121

126 HALHTVLEERK-NLYRHKGNPLPRKASVYARVAGDHVLTVDISTGRVDSNSPV 184

122 LVNG-PLPHBSYMDVVERHGGPLPYHRRPVLFQELVDIDQMFNRELPAASGLHHCC 180

185 RVSGYPMWTIEEMNDIV-----VPSNAD-----FNRTIIS----- 216

181 FYKRGRLVMTAPRGLSGDRATWFG-----LYYNISAGFELHHV-GLE 227

217 ---RGVNLIDVICFP--ISCG---WFGNKEENARVKSOCFMTQGTFRNYMRIDELT 266

228 LTVNHKALDPAKWTIQKVFYQGRYDLSLAOLEAGLVNVVLIIPDNGTGSLSKSPV 287

267 ILID--LDTKQ-----VIEITDGR-----IPI 288

288 PPG-----PAPLQF-YPOGPRFSVQGRVAS-SLTFPSGLGAFSGP 328

289 PGSTNTDYRFQKLAITTDKTRPLNPISIEQPRGPFVIEDNHLVKKANNEFLKPPRAGV 348

329 RLFVDR-----FGGERLYVEISLOALAIYGSNSPAMTFRVYDGG-FGMGKYTTPLR 381

349 VISRRVHDPDTHERRDVTYKGEVSELFVPIYDPEDAWFKTYMAGEYGFGLQMLPLVP 408

382 GVDCCYLAIVYDMHFLLESQAARKTIRDAFCVFEQN-----OGLPLRRHSDI 428

409 LNDCCRNAAIYMDGVRAADGTPFVRENMYCIESYAGDIGMHSSEPIGIRREVR--- 465

429 YSHYFGGLAETVLVYRSMSTLNDYVMDTVFHPGALFIRYALGIYSALFEGATGY 488

466 -----PKVTLVYRMAASVGNNDYIIDYEFOTDGLIKAKGLSGIL---WYKGTYYON 514

489 GNOV-----SEHTLGTVHTSAHFKNVDVAGLEWVAEDMVFYPMAY 532

515-KNOVKKRDKGNEELHGLTISENVIGVINDHTVTFILDLVDGPPD-----SFVKYNLKR 569

533 PMSPEHOLR---LQVTRKLEMEEOAFLVGSATP-RYLYLASNHNKMGPRGRYIOM 588

570 QETEGESPRKSYLAKVARIATKENDQIKLSIDPSEEHVINSKGTIVGNPTGYKVV 629

589 LSFAGEPLPONSMAFGESMERYOLAVTORKEEPESSSVENONDPWAPTVD-FSPFINN 647

630 RTTASLSDHDPQKRGAFITNQIWTYFNKSEMGAGLFTYQSHGDDTLAVMSD--RD 687

648 ETIAGCDLVAVWTAGFLIHPAEDIPNTYVYGCVGFILRPVNFPEDESPFSADSIYR 707

Db 688 RDIENKDIWVMTLGFHHIPQEDFPIPTVSS--SFDLKPVNFFERNILSAAPNF--- 742

QY 708 GQDDAGACEYNPLA 721

Db 743 -EHDLPVCGVQSVS 755

RESULT 11

G71412

probable amine oxidase - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

A:Variety: Columbia

C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 20-Jun-2000

C:Accession: G71412

R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.;

P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weissenegger, T.; Pohl, T.M.; Terry, N.;

avanagh, T.; Hempel, S.; Kotter, P.; Entlian, K.D.; Rieger, M.; Schaeffer, M.; Funk,

Nature 391, 485-488, 1998

A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Pulgido

erhoff, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Reichman, S

C.; Chalwatzis, N.

A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsi:

A:Reference number: A71400; M01D:9812113; PMID:9461215

A:Accession: G71412

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-650 <BEV>

A:Cross-references: GB:297337; NID:q2244829; PID:q2244851

C:Genetics:

A:Map position: 4COP9-4G3845

C:Superfamily: amine oxidase (copper-containing)

Query Match 8.8%; Score 353.5; DB 2; Length 650;

Best Local Similarity 23.9%; Pred. No. 6.3e-20; Indels 139; Gaps 29;

Matches 161; Conservative 96; Mismatches 278;

80 VELDPPKAAALAHLD---DRGSPPPAREALAIFFGRQOPNVSELYVGLPHPSYRD 135

56 LDLEPPKSHLQMLSPKSPKPPRRRSYVVRAGQ---TEELI-----D 101

QY 136 TVVER-----HGGP---LPHRRPVLFQELVDIDQMFNREL--PAASGLHHCC 180

102 LTTSKIASRIYTGHPFSFTFIELEFKASKLPLTPPKKSLDLSLISEVSCIPFTVG 161

QY 181 FYKRGRLVMTAPRGLSGDRATWFGGLVYNISAGFELHHV-GLELVNKKALDPA 239

162 WYGE-----TTTRRELKA-----SCFTRDGSVNVFTPIGITYTID--VD--- 200

240 WTQKVFYQGRYDLSLAOLEAGLVNVVLIIPDNGTGSLSKSPVPPGAPPLQFY- 298

201-SMÖYIKISDRKRP-----IPK-BENDRRTK-----RFPFFC 234

QY 299-PQGRFVSQGRVASSLMTESFGLGAFSGPFIYD-----RQGERLYVEISLOEA 349

223 NVSDTGFKILGRVKKAMKFFHVGFTARAGVITISASYLDPRTKRR--RVMYRGHVST 292

QY 350 LAIYGSNSPAAITTYVD-CGFGMGKYTTPPLRGDCPRLATYVDMHFLLESQAARKTIR 408

293 FVYMDPTYEWYRTFMDIGFEGFGSAVNLPLDPCPNAAFLDGHVAGPGCTQAKMTN 352

QY 409 ACFVEONGLPLRRHSDLY---SHYFGLAETVLVYRSMSTLNDYVMDTVFHPGA 465

Db 353 VNCVEEKN-GYASRPHEDINVPQVITISGEAEISLVYVMVATLGNVYDIYVMEFKKGA 411

QY 466 IEI-----RFYATGYISSAFLEGATKYGNQVSEHTLGVYHTSAHFKNVDVAGLE 517

Db 412 IRVGYDVLGVLEVKATSYSDQI--TENNYGTIVAKNTIAVNHCHYLYLVDLDVNG 469

QY 518 NVYMAEDVYFVMA-VPNAPPEHOLQVOTRKLEMEEOAFLVGSATPFRYLYLASNHN 576

Db 470 NSLVAKKLTIVRTVEYVKNKSSRRKSYWTYVKEAKLEADGRVRLGSDPVELLIVNKKTT 529

QY 577 KMGHPGRYIOMLSFAGEPLP---QNSMAFGESMERYOLAVTORKEEPESSSVENON 633

Db 530 KIGNTVGYRL-----IPEHLQATSLITLDDDDPELRAGYT-----KRPVAVTAYDSE 576
 Qy 634 PWAPVDESD-----FINNETAGKDLVAVTAFGLHIAEDIPNTVTVGNGV 682
 Db 577 RWAGGF-YSDNSRGDDGLAVSSRRREIENKDIWVMYNGFHIIYOEDEFVPTLHG-- 633
 Qy 683 GFELRPYNEFDEDP 696
 Db 634 GFTLRPSNFPNDP 647

RESULT 12

JC2139
 Phenylethylamine oxidase (EC 1.4.3.-) - *Arthrobacter globiformis*
 C:Species: *Arthrobacter globiformis*
 C:Date: 30-Sep-1993 #sequence_revision 20-Aug-1994 #text_change 28-May-1999
 C:Accession: JC2139; PC2070
 R:Tanizawa, K.; Matsuzaki, R.; Shimizu, E.; Yorifuji, T.; Fukui, T.
 Biochem. Biophys. Res. Commun. 199, 1096-1102, 1994
 A:Title: Cloning and sequencing of phenylethylamine oxidase from *Arthrobacter globiformis*
 A:Reference number: JC2139; MUID:94197690; PMID:8147851
 A:Accession: JC2139
 A:Molecule type: DNA
 A:Residues: 1-638 <TAN>
 A:Cross-references: GB:U03517; NID:9451488; PID:AA18114.1; PID:9451489
 A:Accession: PC2070
 A:Molecule type: Protein
 A:Residues: 3-18;38-52;132-142;185-209;219-235;402-419;474-497;504-537;540-559;611-625
 C:Comment: This enzyme catalyzes the oxidative deamination of various biogenic primary
 C:Superfamily: amine oxidase (copper-containing)
 C:Keywords: copper binding; oxidoreductase; quinoprotein; topaquinnone
 F:355,411,433/Binding site: copper (His) #status predicted
 F:382/Modified site: topaquinnone (Tyr) #status predicted

Query Match 8.58; Score 340; DB 2; Length 638;
 Best Local Similarity 23.7%; Pred. No. 7.4e-19;
 Matches 157; Conservative 100; Mismatches 274; Indels 132; Gaps 28;

Qy 87 KAALAHLD--RGSPPPAREALAIYF---GROPQNVSELVGPLHPSMDVTER 140
 Db 38 RIALVGLVDPARGAGSEMEDRRFRFIDVSGARQ-----EVTYSV 79
 Qy 141 HGGPL-----PYHRRPVLFQEYLDIDIMFN-----RELPOAS 173
 Db 80 TNGVYISAVELDTAATGELPVLEEFVEVQDLATDERMLKALARNLDVSKVRAPLSA 139
 Qy 174 GLLHHCCTYKRGRLVMTAPRGL---QSGRATYFGLYINISGAGFLHVGLEL 229
 Db 140 GVFEVA--EERGRRL-----RGLAFVQDPEDPSAW--AHVYDGLVAVYDVYSKEVT 187
 Qy 230 VNHKALDPARTIQKVFQGRYVDSLAQLEAFAGLVNVVLPDNGSGWSLSKSPRP 289
 Db 188 ---KVIDGVFVPA--EHGNTDP-----ELTGPLRT 215
 Qy 290 GRAPPLQFYPOGPRFSVO--GSRVASSLMTFSGLGAFSGPRIFDVRFQ--GERL---VVEI 344
 Db 216 TQKDISITQPGSPFTVGNHIEWEKMSLDVGFVGRGVVLAHNAFQDGRRLRIINRA 275
 Qy 345 SLOGALAIYGNNSPAAATTKYVD--GGFGMKYITPLTRGVDC---PYLATVYDMHFLLE 399
 Db 276 SIAMVVPYGPSPRISQWNTFDGEYLVGQYANSLGCGCLDITLSPVISAQF--- 332
 Qy 400 SQAKTIRDACVFEONGLPLRRHSDLYSHYFGGLAEVLYVRSMTLLNDVYMDTV 459
 Db 333 -GNREIRNGICMHEEDMGLIAK--HSDLMGICITRRNRKRVISFTTIGVYDGGFWY 389
 Qy 460 FHPGGAIEIRFYATGYI--SSAFLEGAQKYGXQVSEHTLGTVHSHAFKVDLVAGLEN 518
 Db 390 LYLDGITEFEAKATGVFTSAFPEGSDNI--SQLAPGLGAFPHQHIFGARLDMAIDGETN 448
 Qy 519 WMAEDVVFVMAVPSPEHQLQVTRKLLMEEQAAFLVGSATPPTLYLASHS--NK 577

Db 449 RVEEDVVRQTMCPGNERGNMFSR---KRTVLTRESEAVREADARTGRTWITSNPSKNR 505
 Qy 578 WCHPRGRI---QMSFAGEPLPONSSMAKGFSEWERYOLAVTORREEPSSSVF--NND 633
 Db 506 INEPVGYKLHAHQPTLLADP---GSSIAARRAAFKTKLMTWRYADDERYPFGDVNGHS 562
 Qy 634 PWAPVDESDFINNETAGKDLVAVTAFGLHIAEDIPNTVTVGNGVFLRPYNEFD 693
 Db 563 GGAGLPST--IAQDRIDGODIVVMTFTGLTHPRVEDMP--IMPVDVGFELRPEGFD 618
 Qy 694 EDP 696
 Db 619 RSP 621

RESULT 13

JC7251
 amine oxidase (copper-containing) (EC 1.4.3.6) - garden pea
 C:Species: *Pisum sativum* (garden pea)
 C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 21-Jul-2000
 C:Accession: JC7251
 R:Koyanagi, T.; Matsumura, K.; Kuroda, S.; Tanizawa, K.
 Biosci. Biotechnol. Biochem. 64, 717-722, 2000
 A:Title: Molecular cloning and heterologous expression of pea seedling copper amine o
 A:Reference number: JC7251
 A:Accession: JC7251
 A:Molecule type: mRNA
 A:Residues: 1-674 <KOY>
 A:Cross-references: DDBJ:AB026253
 A:Experimental source: seed
 C:Comment: This enzyme, a homodimer, containing one Cu²⁺ ion and one 2,4,5-trihydroxy
 C:Superfamily: amine oxidase (copper-containing)
 C:Keywords: copper; copper binding; homodimer; oxidoreductase

Query Match 8.18; Score 325.5; DB 2; Length 674;
 Best Local Similarity 24.18; Pred. No. 1.1e-17;
 Matches 146; Conservative 85; Mismatches 233; Indels 141; Gaps 30;

Qy 158 LDIDOMITNRELPOASGLHHCCT---YKRGRLVMTAPRGLQSGDRATMGLYINI 214
 Db 128 LSVDEQSLAIEIP---LKYPPFIDSVYKRGRLNSEIYC-----SSFYMGWGEKKNV 176
 Qy 215 -----SGAGFFLHNY--GLELVNHNKALDPARTIQKVFQGRYVDSLAQLEAF 262
 Db 177 RTVRLDCMKSESTYNIYRPTGITVAD---LD-----LKKIV---EYHNR----- 217
 Qy 263 EAGLVNVVLPD--NGTGSWSLSKSPVPPAPR--LQFYPOGPRFSVQSGSRVASSLMTFS 319
 Db 218 -----DIAVPTAEKTEVQVSKSP--PRGRKQNSLTSHPQGRGQIEGHSVANMKFH 271
 Qy 320 FGLGAFSG-----PRIFPV--RFGGERLYEISLOBALAIYGNNSPAAATTKRYVDG--FGM 372
 Db 272 IGFVFRAGIVISLAIYDLEKHSRVLVYKGYISELFVYQDPTPEEFYKFFPDSGEGF 331
 Qy 373 GKTYTPLRGVDCPYLATVYDMHFLLESQAPKTRIDAFCEVQONGLPLRRHSDLYSHY 432
 Db 332 GLSTVSLIPNDCRPHAQFIDTTHSANGPILKNAICVEQ-----YGINMMHT 383
 Qy 433 FGL-----AEVLYVRSMTLLNDVYMDTVFHPGSAIEIRFYATGYISSAFLEG 483
 Db 384 ENGIPNESIESRFEVNLIVRTIYGVGVNDVIMEFASSINPALASGILE---IKG 440
 Qy 484 ATGK-----YGNVSEHTLGTVHSHAFKVDLVAGLENWMAEDVVFVPM--- 530
 Db 441 TNIRKDEIKEDLGKLVANSISGIYHDFYIYLDPIDIDTSHSFEKTSIKTVRIKGS 500
 Qy 531 ---AVMPSHQLOLQVTRKLLMEEQAAFLVGSAPRYLYLASNSNKKGHPGRYRQ 587
 Db 501 SKRSYSTTEQTAK-----TESDAKTTIGLAPRELVVVPMNTKAVGNEVGRLL- 550
 Qy 588 MLSFAGEPL-----PONSSMARGFSERYOLAVTORKEEPSSSVFNQNDPMAP--TV 639

Db 551 IPAIRAHPLFEDDYPO-----IRG-AFTNYNWT-----AYNRETKMAGLXY 594
Qy 640 DFS---DFT-----NNETAGKDYAVMTAGFLHIAHADIPTVTVGNGVFLRPYNF 691
Db 595 DHSRCDTLAVMTKONREIVNKDIYMHVYGHVPAQDFLIMPL--STSELRPTNF 652
Qy 692 FEEDP 696
Db 653 FERNP 657

RESULT 14
747403
amine oxidase-like protein - Arabidopsis thaliana
N:Alternate names: protein F23N14.50
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 21-Jul-2000
C:Accession: T47403
R:Blocker, H.; Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.; Quetier, F.; Salanoubat
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24465
A:Accession: T47403
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-687 <BIO>
A:Cross-references: EMBL:AL138638
A:Experimental source: cultivar Columbia; BAC clone F23N14
C:Genetics:
A:Map position: 3
A:Intons: 145/3; 303/1; 402/3; 440/3; 591/3; 624/2
A:Note: F23N14.50
C:Superfamily: amine oxidase (copper-containing)

Query Match 8.1%; Score 324; DB 2; Length 687;
Best Local Similarity 22.1%; Pred. No. 1.6e-17;
Matches 168; Conservative 102; Mismatches 298; Indels 192; Gaps 31;

Qy 20 PHCSVSAPQPTWHPQSQSLFADLSBELTAVNRFLTQRGLVDAQAARPDNCVFS 79
Db 39 PHNP-----LDPLTTP-----EIKRVOTILSGH-DPFGSGS-----TIHA 74
Qy 80 VELDLPKAAALAHLDGSPAREALAYFFGRQOPNV---SELVVGPLPHPSY----- 132
Db 75 MALDEPKORYIRKKKGRDLPRRAELIAMSNGSHVLTDLKGRVYSDLVNTPFGPI 134
Qy 133 --MRDVTVERHGRLPYHRRPVEFOEYLDIDOMIFNRELPOASGLLHHCFCYKRRGNLV 190
Db 135 LTMMDIAY---SQVPY-----KSYEFNRISIEARGIP-FSGLICITPPAGWGPDP-- 180
Qy 191 TMTTAPRGLSGDRATWFGLYYINISGAGFLHHV-GLELLVNHKALDARMTIQVFG 249
Db 181 -----EGRARYIKQCFSKODTVNFMPIEGLVLTDMKLE----- 218
Qy 250 RYDLSLAQLEQFEAGLVNVLLPDNGGSGMSLSPVPPGAPLPQ----- 297
Db 219 -----IKIYDNG-----PVYPKSGTGTFRXGFLNETVYMDRV 252
Qy 298 -----YPOGRFVSVO-GSRVASSLWTFSGLGAFSGRIFDVROGERLYEISLEAL 350
Db 253 NPMSEOPDGRSFGVEDGYLVKMAWKFHIFPDORAGMI-----SOATVADSKTGEAR 306
Qy 351 AI-YGNSPAM-----TTRYVDG-GFMGKXTTPLTRGVCOPYLATYVDMHFL 397
Db 307 SVMKGFASELFVPMNMDGEGYSKAYMDAGEFGSPSMLVPLINDCPRAAYITDGFPA 366
Qy 398 LESQAPKTRIDAFVCEONOGPLRRHSDLSHYFGGL-----AETLVVRSKSTLN 451
Db 367 SPEGIPILOPNMNICFERYAGDTSWRHSEIL-----LPQVDIRESAKYTLVARMASGVN 422
Qy 452 YDYVMDTVFHPSGAIEIRFYATG-----YISAFLEGATGKGNNOVSEHTLGTVTH 503
Db 423 YDIIDMEFQMDGVIRVTVAASGMILVKGATAYENVEDLGEKEDSDGPLLSENVIGVVDH 482

Qy 504 SAHFVLDVAGLENNWMAEDNVFVMAVPWSPEHOLR---LQVTRKLEMEQAAFLV 560
Db 483 FISFHLMDIDDSAN-----NSFVKVHLERKQRLPPGSRRSKYLKVKYAKTEKDAQIRM 538
Qy 561 GSATPRYLTLAS-NHSNKGHRPGYRIOMLSFAGEPLPONSASARFSEMEROLATYORK 619
Db 539 SLVDPYEFHLVDPNRLSRGNPAGYKLVPGNNAASLLDDHDPQMRGATFNQIWTYR-- 596
Qy 620 EEEPPSSSVFNONDPWA-----PTVD-FSDFINNETIAGKDLVAMTAGFLHIP 667
Db 597 -----YNRSEQWAGGLIMTQSGEDITLQVMSD--NRSTENNDIYMTLGTGHHNP 645

Qy 668 HAEDIPNTVTVGNGVGFLLRPYNFEDDP-----SFYSAD 702
Db 646 QOEDFPVMTIAS--SFELKPVNFESNPVLG:SPFEKD 683

RESULT 15
C44239
amine oxidase (copper-containing) (EC 1.4.3.6) precursor - garden pea
C:Species: Pisum sativum (garden pea)
C>Date: 31-Dec-1993 #sequence_revision 05-Jan-1996 #text_change 17-Mar-2000
C:Accession: A57327; C44239
C:Key words: A.J.; McPherson, M.J.
J. Biol. Chem. 270, 16939-16946, 1995
A:Title: Cloning and molecular analysis of the pea seedling copper amine oxidase.
A:Reference number: A57327; PMID:95348126; PMID:7622512
A:Accession: A57327
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-674 <TIP>
A:Cross-references: GB:L39931; NID:9685197; PID:9685198
R:Jones, S.M.; Palcic, M.M.; Scamam, C.H.; Smith, A.J.; Brown, D.E.; Dooley, D.M.;
Biochemistry 31, 12147-12154, 1992
A:Title: Identification of topaquinone and its consensus sequence in copper amine
A:Reference number: A44239; PMID:93090748; PMID:11457410
A:Accession: C44239
A:Molecule type: Protein
A:Residues: 409-417, 'X', 419 <JAN>
A:Experimental source: seedling
C:Superfamily: amine oxidase (copper-containing); oxidoreductase; quinoprotein; tol
C:Keywords: copper; glycoprotein; metalloprotein; status predicted <SIG>
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-674/Product: amine oxidase (copper-containing) #status predicted <MAT>
F:166,389,583/Binding site: carbonyl site (Asn) (covalent) #status predicted
F:382,467,469/Binding site: copper (His) #status predicted
F:412/Modified site: topaquinone (Tyr) #status experimental

Query Match 8.0%; Score 320; DB 2; Length 674;
Best Local Similarity 25.8%; Pred. No. 3.2e-17;
Matches 120; Conservative 66; Mismatches 189; Indels 90; Gaps 20;

Qy 282 SLKSEVPFGPAPP--LQETPGGPRSVSGSRVASSLWTFSGLAFSG-----PRIPEV- 333
Db 223 SKQSP-PRGPKRQHSLSHQPOGPGQINGHSVSMANMFHIGEDVRAQIVISLASTYDE 291
Qy 334 RFOGERLYEISLOEALAYGNSPAAATTTRYVDG-GFMGKXTTPLTRGVCOPYLATYV 392
Db 292 KHKSRVLYKGYISLEFLPYDPTIEFFYKFTFDGEGFGGLSTYSLIPNRCPHQAQFI 351
Qy 393 DMHFLLESQAPKTRIDAFVCEONOGPLRRHSDLSHYFGGL-----AETLVVY 443
Db 352 DTIVHASANGTPIILKNALCVFQ-----YGNIMMRHTENGIPNESIESRTEVNLI 403
Qy 444 RSMSTLLTYDVMVDFVHPSGAIEIRFYATGYISSAFLGAGK-----YGNVSE 494
Db 404 RTIVTVGNDVYIDMEFASGSGIKPSIALSGILE---IKGTINKRDEIKEDLHGLKLYSA 460
Qy 495 HTLGFVHTHSAHFKVLDVAGLENNWMAEDNVFVPM-----AVWSPPEHOLQRLQYTR 547
Db 461 NSIGITYHDHFTYTYDIDGIDGTHNSFEKSLTVRKDQSSKRSKRWTEGTQAK----- 515
Qy 548 KLEMEQAAFLVGSATPRYLTLASNHSNKGHRPGYRIOMLSFAGEPL-----PONS 601


```
Db 516 ----TESDAKITIGLAPALVYVNNIKTAVONEVGIRL-IPAIPAHLITEDDYPO--- 567
OY 602 MARGESMERIOLAVTQKKEEPPSSSVFNQNDPMAP--TVDES---DFI-----NNETIA 651
Db 568 -IRG-AFTNYVWVY7-----AYNRTEKNAGGLYVDHSRGDDTLAVVTKQNRREIV 614
OY 652 GKDLVAVWTAGFLHPHEDIPNTVTGNGVGFILRPYNFDEDP 696
Db 615 NKDVMHMHVGIHHVPAOEDFPIMPL--STSFELRPTNFFERNP 657
```

Search completed: May 20, 2003, 12:05:04
Job time : 23 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:41:09 ; Search time 17 Seconds
(without alignments)
1815.199 Million cell updates/sec

Title: US-10-081-408-2_COPY_20_763
Perfect score: 4010
Sequence: 1 LVCVLLVGRGDGGEPSQLP.....QNAACAPDLPAFSGGFSHN 744

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4010	100.0	763	1 AOC3_HUMAN	016853 homo sapien
2	3404	84.9	765	1 AOC3_MOUSE	070423 mus musculu
3	3343	83.4	762	1 AOC3_BOVIN	046406 bos taurus
4	3331	83.1	762	1 AOC3_BOVIN	029437 bos taurus
5	2565	64.0	756	1 AOC2_HUMAN	075106 homo sapien
6	1398.5	34.9	746	1 ABR_RAT	P36633 rattus norv
7	1394	34.8	751	1 ABR_HUMAN	P19801 homo sapien
8	1354	33.8	751	1 AOC3_RAT	008590 rattus norv
9	413	10.3	755	1 AOC3_KLEAE	P49250 klebsiella
10	397.5	9.9	757	1 AOC3_HUMAN	P46883 escherichia
11	377	9.4	683	1 AOC3_HUMAN	059118 arthrobacte
12	377	9.4	683	1 AOC3_HUMAN	007121 arthrobacte
13	363	9.1	648	1 AOC3_HUMAN	007123 arthrobacte
14	363	9.1	648	1 AOC3_HUMAN	P46881 arthrobacte
15	340	8.5	638	1 AOC3_HUMAN	P46881 arthrobacte
16	320	8.0	674	1 AOC3_HUMAN	P43077 plasm sativ
17	310	7.8	666	1 AOC3_HUMAN	P43077 plasm sativ
18	295	7.4	671	1 AOC3_HUMAN	P43077 plasm sativ
19	111	2.8	692	1 AOC3_HUMAN	P12807 plicella fas
20	103	2.6	2224	1 AOC3_HUMAN	090416 plicella fas
21	102.5	2.6	591	1 AOC3_HUMAN	P12859 homo sapien
22	101.5	2.5	736	1 AOC3_HUMAN	P98000 bradyrhizob
23	101.5	2.5	1520	1 AOC3_HUMAN	060838 mus musculu
24	101	2.5	435	1 AOC3_HUMAN	046837 escherichia
25	100.5	2.5	1237	1 AOC3_HUMAN	054804 mus musculu
26	100	2.5	736	1 AOC3_HUMAN	P32004 homo sapien
27	98.5	2.5	1217	1 AOC3_HUMAN	014641 homo sapien
28	98.5	2.5	1217	1 AOC3_HUMAN	P39665 phaseolus v
29	98	2.4	2212	1 AOC3_HUMAN	005193 homo sapien
30	97	2.4	733	1 AOC3_HUMAN	005198 ebola virus
31	96.5	2.4	1137	1 AOC3_HUMAN	P17127 blaberus di
32	96	2.4	421	1 AOC3_HUMAN	P08543 herpes simp
33	95.5	2.4	553	1 AOC3_HUMAN	P34945 thermus the
					023066 arabidopsis

34	95.5	2.4	985	1 AGU1_ASPIRG	P56526 aspergillus
35	95.5	2.4	2175	1 POLG_BOVEY	P12315 bovine ente
36	95	2.4	2258	1 PA5_PIG	0991P1 sus scrofa
37	94.5	2.4	901	1 VEF_GVP	P41723 pseudolatia
38	94.5	2.4	1522	1 PST1_SCHPO	009750 schizosacch
39	94.5	2.4	2555	1 PPS3_BACSU	P39847 bacillus su
40	94	2.3	492	1 SYM_YEAST	P48527 saccharomyc
41	94	2.3	716	1 DVL3_HUMAN	092897 homo sapien
42	94	2.3	716	1 DVL3_MOUSE	061062 mus musculu
43	93.5	2.3	878	1 FIMD_ECOLI	P30130 escherichia
44	93	2.3	697	1 YN26_MYCTU	P71886 mycobacteri
45	93	2.3	958	1 MML1_MYCTU	P95211 mycobacteri

ALIGNMENTS

```

RESULT 1
AOC3_HUMAN          STANDARD:      PRT:      763 AA.
AC  016853;
DT  01-NOV-1997 (rel. 35, Created)
DT  01-NOV-1997 (rel. 35, Last sequence update)
DT  15-JUN-2002 (rel. 41, Last annotation update)
DE  Membrane copper amine oxidase (EC 1.4.3.6) (Vascular adhesion protein-
DE  1) (VAP-1) (HPAO).
GN  AOC3 OR VAP1.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX  NCBI_Taxid=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Placenta;
RX  MEDLINE=97128319; PubMed=8972912;
RA  Zhang X., McIntire W.S.;
RT  *Cloning and sequencing of a copper-containing, topaquinine-
RT  containing monamine oxidase from human placenta.*;
RL  Gene 179:279-286(1996).
RN  [2]
RP  SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC  TISSUE=Lung;
RX  MEDLINE=98317014; PubMed=9653080;
RA  Smith D.J., Salini M., Bono P., Hellman J., Leu T., Jalakanen S.;
RT  *Cloning of vascular adhesion protein 1 reveals a novel
RT  multifunctional adhesion molecule.*;
RL  J. Exp. Med. 188:17-27(1998).
CC  - FUNCTION: CELL ADHESION PROTEIN THAT PARTICIPATE IN LYMPHOCYTE
CC  RECIRCULATION BY MEDIATING THE BINDING OF LYMPHOCYTES TO
CC  PERIPHERAL LYMPH NODE VASCULAR ENDOTHELIAL CELLS IN AN L-SELECTIN-
CC  INDEPENDENT FASHION. HAS A MONAMINE OXIDASE ACTIVITY.
CC  - CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(2) +
CC  H(2)O(2).
CC  - COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
CC  - SUBUNIT: HOMODIMER.
CC  - SUBCELLULAR LOCATION: Type II membrane protein.
CC  - TISSUE SPECIFICITY: MOST STRONGLY EXPRESSED ON THE HIGH
CC  ENDOTHELIAL VENULES OF PERIPHERAL LYMPH NODES AND ON HEPATIC
CC  ENDOTHELIAL.
CC  - PTM: Topaquinine (TPQ) is generated by copper-dependent
CC  autooxidation of a specific tyrosyl residue (By similarity).
CC  - PTM: N- AND O-GLYCOSYLATED.
CC  - SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC  - SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC  This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@sib-sib.ch).
CC  EMBL: U39447; AAC50919.1; -

```

```

DR EMBL: AF067406; AAC25170.1; -.
DR Genem; HGNC:550; AOC3.
DR MIM: 603735; -.
DR InterPro: IPR000269; CUNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
DR Pfam: PF02727; Cu_amine_oxidn3; 1.
DR Pfam: PF02728; Cu_amine_oxidn3; 1.
DR PRINTS: PRO0766; CUDAOXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
DR Signal: Signal: Copper; TPO; Glycoprotein; Transmembrane;
KW Signal-anchor; Cell adhesion; Polymorphism; Metal-binding.
FT DOMAIN 1 5
FT TRANSMEM 6 26
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 27 763
FT MOD_RES 471 471 TOPAQUINONE (BY SIMILARITY).
FT METAL 520 520 COPPER (POTENTIAL).
FT METAL 522 522 COPPER (POTENTIAL).
FT METAL 684 684 COPPER (POTENTIAL).
FT BINDING 673 673 AMILORIDE (BY SIMILARITY).
FT CARBOHYD 137 137 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 232 232 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 294 294 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 592 592 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 618 618 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 666 666 N-LINKED (GLCNAC... ) (POTENTIAL).
FT VARIANT 317 317 Y -> H (IN DBSNP:435287).
FT /FTID=VAR_012064.
SQ SEQUENCE 763 AA; 84621 MW; 58AD55605EC9D228 CRC64;

Query Match 100.0%; Score 4010; DB 1; Length 763;
Best Local Similarity 100.0%; Pred. No. 3,1e-307;
Matches 744; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LVCLVLRGSGDGEPSQLPHCPSPVSAQPTWHPGOSQLFADLSREELAVMFLTQRL 60
DB 20 LVCLVLRGSGDGEPSQLPHCPSPVSAQPTWHPGOSQLFADLSREELAVMFLTQRL 79
QY 61 GPGVLDAAQAPSPDNCVFSVELQPPKAAALAHDRGSPPARRELAIVFGROPQPVNS 120
DB 80 GPGVLDAAQAPSPDNCVFSVELQPPKAAALAHDRGSPPARRELAIVFGROPQPVNS 139
QY 121 ELVVGPLPHPSYMDVVERHGGLPPIHRRPVLFOEYDIDIDOMIFNRELPAQSGLLHHC 180
DB 140 ELVVGPLPHPSYMDVVERHGGLPPIHRRPVLFOEYDIDIDOMIFNRELPAQSGLLHHC 199
QY 181 FYKHGRNLVTMTAPRGLOSGDRATWGLYNNISGAGFLHVGLELLVNHKALDPAWM 240
DB 200 FYKHGRNLVTMTAPRGLOSGDRATWGLYNNISGAGFLHVGLELLVNHKALDPAWM 259
QY 241 TIQVFTQGRYYDSLAEQFEAGLVNVLIPDNGTGSWSLSPVPPGAPPLQTFPQ 300
DB 260 TIQVFTQGRYYDSLAEQFEAGLVNVLIPDNGTGSWSLSPVPPGAPPLQTFPQ 319
QY 301 GPRSVOSRAVSLMPSFGLAFSGPRIFDVROGRIYVEISLOALAIYGSNPA 360
DB 320 GPRSVOSRAVSLMPSFGLAFSGPRIFDVROGRIYVEISLOALAIYGSNPA 379
QY 361 MTRRYVDGFGMGKYYTTPPLRGVDCPYLATVYDMHFLLESQAPKTRIRAFCEFNOCILP 420
DB 380 MTRRYVDGFGMGKYYTTPPLRGVDCPYLATVYDMHFLLESQAPKTRIRAFCEFNOCILP 439
QY 421 LRRHSDLSYHFGGLATVVLVNSMSTLNYDYWDVVFHPSGAIETRFATGYSNAF 480
DB 440 LRRHSDLSYHFGGLATVVLVNSMSTLNYDYWDVVFHPSGAIETRFATGYSNAF 499
QY 481 LFGATGKGNVSEHTLTCTVHTSHAFKVDVAGLEMMVNAEDVAFPMAPVAPSPHQ 540
DB 500 LFGATGKGNVSEHTLTCTVHTSHAFKVDVAGLEMMVNAEDVAFPMAPVAPSPHQ 559
QY 541 ORLOVTRKLEMEQAAFLVGSATPRYLVLASNSNKGHPRGYRIOMLSFAGEPLPONS 600
DB 560 ORLOVTRKLEMEQAAFLVGSATPRYLVLASNSNKGHPRGYRIOMLSFAGEPLPONS 619
QY 601 SMARGSMERYQLAVYORKKEEPPSSSVFQNDNPAPVDFSPFINNETIAGDLYAMT 660
DB 620 SMARGSMERYQLAVYORKKEEPPSSSVFQNDNPAPVDFSPFINNETIAGDLYAMT 679
QY 661 AGFLHPIHADIPNTVYVGVGFELRPYNFEDDPSEFYSAISYFGDDADACEVNP 720
DB 680 AGFLHPIHADIPNTVYVGVGFELRPYNFEDDPSEFYSAISYFGDDADACEVNP 739
QY 721 ACPLQAAACAPDLPAPFSGHGFSHN 744
DB 740 ACPLQAAACAPDLPAPFSGHGFSHN 763

```

```

RESULT 2
AOC3_MOUSE
ID AOC3_MOUSE STANDARD; PRT; 765 AA.
AC 070423;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Membrane copper amine oxidase (EC 1.4.3.6) (Vascular adhesion protein-1) (VAP-1).
DE 1) (VAP-1).
GN AOC3 OR VAP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid:10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C, and 129/SvJ;
RX MEDLINE:98414290; PubMed:9743358;
RA Bono P., Salmi M., Smith D.J., Leppanen I., Horelli-Kultunen N.,
RA Palotie A., Jalkanen S.;
RT Isolation, structural characterization, and chromosomal mapping of
RT the mouse vascular adhesion protein-1 gene and promoter."
RL J. Immunol. 161:2953-2960(1998).
CC -1- FUNCTION: CELL ADHESION PROTEIN THAT PARTICIPATE IN LYMPHOCYTE
CC RECIRCULATION BY MEDIATING THE BINDING OF LYMPHOCYTES TO
CC PERIPHERAL LYMPH NODE VASCULAR ENDOTHELIAL CELLS IN AN L-SELECTIN-
CC INDEPENDENT FASHION. HAS A MONOMINE OXIDASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +
CC H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topequinone per subunit.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
CC -1- PTM: Topaquione (TPO) is generated by copper-dependent
CC autoxidation of a specific tyrosyl residue (by similarity).
CC -1- PTM: N- AND O-GLYCOSYLATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF054831; AAC3747.1; -.
DR EMBL: AF078705; AAC35839.1; -.
DR MGD: MGI:1306797; AOC3.
DR InterPro: IPR000269; CUNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
DR Pfam: PF02727; Cu_amine_oxidn3; 1.
DR Pfam: PF02728; Cu_amine_oxidn3; 1.
DR PRINTS: PRO0766; CUDAOXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
DR Oxidoreductase; Copper; TPO; Glycoprotein; Transmembrane;
KW Signal-anchor; Cell adhesion; Metal-binding.
FT DOMAIN 1 6
FT CYTOPLASMIC (POTENTIAL).

```

Query Match
Best Local Similarity 83.9%; Score 3404; DB 1; Length 765;
Matches 618; Conservative 61; Mismatches 64; Indels 0; Gaps 0;

FT	TRANSMEM	7	27	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT	DOMAIN	28	765	(POTENTIAL).
FT	MOD_RES	471	471	EXTRACELLULAR (POTENTIAL).
FT	METAL	520	520	TOPOQUINONE (BY SIMILARITY).
FT	METAL	522	522	COPPER (POTENTIAL).
FT	METAL	522	522	COPPER (POTENTIAL).
FT	METAL	522	522	COPPER (POTENTIAL).
FT	BINDING	673	673	AMILORIDE (BY SIMILARITY).
FT	CARBOHYD	137	137	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	232	232	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	294	294	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	592	592	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	659	659	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	666	666	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SEQUENCE	765 AA;	84533 MW;	74896D673DBB44D CRC64;

Query Match
Best Local Similarity 83.2%; Score 3404; DB 1; Length 765;
Matches 618; Conservative 61; Mismatches 64; Indels 0; Gaps 0;

QY	1	LVCVLLVGRGDSGEPSQLPCHPSVSPSAQPTWHPGQSOLFADLSREELTAVMKFLTQRL
DB <td>20 <td>LVCVLLVGRGDSGEPSQLPCHPSVSPSAQPTWHPGQSOLFADLSREELTAVMKFLTQRL</td> </td>	20 <td>LVCVLLVGRGDSGEPSQLPCHPSVSPSAQPTWHPGQSOLFADLSREELTAVMKFLTQRL</td>	LVCVLLVGRGDSGEPSQLPCHPSVSPSAQPTWHPGQSOLFADLSREELTAVMKFLTQRL
QY <td>61 <td>GPELVDAQAARPSDNCVFSYELQLPKAAALAHLDKSGPPAREBALATVFFGRQPPQNV</td> </td>	61 <td>GPELVDAQAARPSDNCVFSYELQLPKAAALAHLDKSGPPAREBALATVFFGRQPPQNV</td>	GPELVDAQAARPSDNCVFSYELQLPKAAALAHLDKSGPPAREBALATVFFGRQPPQNV
DB <td>80 <td>GPELVDAQAARPSDNCVFSYELQLPKAAALAHLDKSGPPAREBALATVFFGRQPPQNV</td> </td>	80 <td>GPELVDAQAARPSDNCVFSYELQLPKAAALAHLDKSGPPAREBALATVFFGRQPPQNV</td>	GPELVDAQAARPSDNCVFSYELQLPKAAALAHLDKSGPPAREBALATVFFGRQPPQNV
QY <td>121 <td>ELVYVGLPSPSYMDVTVESHGGRPLPRHRRPVLFQYELIDDMIFNELFOASGLHHC</td> </td>	121 <td>ELVYVGLPSPSYMDVTVESHGGRPLPRHRRPVLFQYELIDDMIFNELFOASGLHHC</td>	ELVYVGLPSPSYMDVTVESHGGRPLPRHRRPVLFQYELIDDMIFNELFOASGLHHC
DB <td>140 <td>ELVYVGLPSPSYMDVTVESHGGRPLPRHRRPVLFQYELIDDMIFNELFOASGLHHC</td> </td>	140 <td>ELVYVGLPSPSYMDVTVESHGGRPLPRHRRPVLFQYELIDDMIFNELFOASGLHHC</td>	ELVYVGLPSPSYMDVTVESHGGRPLPRHRRPVLFQYELIDDMIFNELFOASGLHHC
QY <td>181 <td>FYKHGRNLTMTTAPRGLOSGDRATWFGLYNLSGAFELHNVGELLYNHNKALDPARM</td> </td>	181 <td>FYKHGRNLTMTTAPRGLOSGDRATWFGLYNLSGAFELHNVGELLYNHNKALDPARM</td>	FYKHGRNLTMTTAPRGLOSGDRATWFGLYNLSGAFELHNVGELLYNHNKALDPARM
DB <td>200 <td>FYKHGRNLTMTTAPRGLOSGDRATWFGLYNLSGAFELHNVGELLYNHNKALDPARM</td> </td>	200 <td>FYKHGRNLTMTTAPRGLOSGDRATWFGLYNLSGAFELHNVGELLYNHNKALDPARM</td>	FYKHGRNLTMTTAPRGLOSGDRATWFGLYNLSGAFELHNVGELLYNHNKALDPARM
QY <td>241 <td>TICKVYVQGRYVDSLAQLEQFAGLVNVLIDNKGSGWSLSPRPAPAPLPQYPO</td> </td>	241 <td>TICKVYVQGRYVDSLAQLEQFAGLVNVLIDNKGSGWSLSPRPAPAPLPQYPO</td>	TICKVYVQGRYVDSLAQLEQFAGLVNVLIDNKGSGWSLSPRPAPAPLPQYPO
DB <td>260 <td>TICKVYVQGRYVDSLAQLEQFAGLVNVLIDNKGSGWSLSPRPAPAPLPQYPO</td> </td>	260 <td>TICKVYVQGRYVDSLAQLEQFAGLVNVLIDNKGSGWSLSPRPAPAPLPQYPO</td>	TICKVYVQGRYVDSLAQLEQFAGLVNVLIDNKGSGWSLSPRPAPAPLPQYPO
QY <td>301 <td>GPRFSVQSGVSSSLMFAFSGCLAFSGPRIFDIPOGERAYELSVQELATLKGNSPAS</td> </td>	301 <td>GPRFSVQSGVSSSLMFAFSGCLAFSGPRIFDIPOGERAYELSVQELATLKGNSPAS</td>	GPRFSVQSGVSSSLMFAFSGCLAFSGPRIFDIPOGERAYELSVQELATLKGNSPAS
DB <td>320 <td>GPRFSVQSGVSSSLMFAFSGCLAFSGPRIFDIPOGERAYELSVQELATLKGNSPAS</td> </td>	320 <td>GPRFSVQSGVSSSLMFAFSGCLAFSGPRIFDIPOGERAYELSVQELATLKGNSPAS</td>	GPRFSVQSGVSSSLMFAFSGCLAFSGPRIFDIPOGERAYELSVQELATLKGNSPAS
QY <td>361 <td>MTRTRYVDSGFGMKYTPPLTRGVDCPYLATYVDMHFLLESQAPKTRDAFCVEQNGLP</td> </td>	361 <td>MTRTRYVDSGFGMKYTPPLTRGVDCPYLATYVDMHFLLESQAPKTRDAFCVEQNGLP</td>	MTRTRYVDSGFGMKYTPPLTRGVDCPYLATYVDMHFLLESQAPKTRDAFCVEQNGLP
DB <td>380 <td>MTRTRYVDSGFGMKYTPPLTRGVDCPYLATYVDMHFLLESQAPKTRDAFCVEQNGLP</td> </td>	380 <td>MTRTRYVDSGFGMKYTPPLTRGVDCPYLATYVDMHFLLESQAPKTRDAFCVEQNGLP</td>	MTRTRYVDSGFGMKYTPPLTRGVDCPYLATYVDMHFLLESQAPKTRDAFCVEQNGLP
QY <td>421 <td>LRRHNSDLYSHYFGGLAETLVYRSMSTLNYDMVDMTHPSGALIRRYATCYISSAF</td> </td>	421 <td>LRRHNSDLYSHYFGGLAETLVYRSMSTLNYDMVDMTHPSGALIRRYATCYISSAF</td>	LRRHNSDLYSHYFGGLAETLVYRSMSTLNYDMVDMTHPSGALIRRYATCYISSAF
DB <td>440 <td>LRRHNSDLYSHYFGGLAETLVYRSMSTLNYDMVDMTHPSGALIRRYATCYISSAF</td> </td>	440 <td>LRRHNSDLYSHYFGGLAETLVYRSMSTLNYDMVDMTHPSGALIRRYATCYISSAF</td>	LRRHNSDLYSHYFGGLAETLVYRSMSTLNYDMVDMTHPSGALIRRYATCYISSAF
QY <td>481 <td>LFQAGKYGVNOSEHTLGTVHTSHARKVDLDVAGLENNVMAEDMVEVPVAPSPSHOC</td> </td>	481 <td>LFQAGKYGVNOSEHTLGTVHTSHARKVDLDVAGLENNVMAEDMVEVPVAPSPSHOC</td>	LFQAGKYGVNOSEHTLGTVHTSHARKVDLDVAGLENNVMAEDMVEVPVAPSPSHOC
DB <td>500 <td>LFQAGKYGVNOSEHTLGTVHTSHARKVDLDVAGLENNVMAEDMVEVPVAPSPSHOC</td> </td>	500 <td>LFQAGKYGVNOSEHTLGTVHTSHARKVDLDVAGLENNVMAEDMVEVPVAPSPSHOC</td>	LFQAGKYGVNOSEHTLGTVHTSHARKVDLDVAGLENNVMAEDMVEVPVAPSPSHOC
QY <td>541 <td>QRLQVTRKLLLEBQAAFLVGSATPRYLTLASNSHNKMGHPRGYRIOMLSFAGEPLDONS</td> </td>	541 <td>QRLQVTRKLLLEBQAAFLVGSATPRYLTLASNSHNKMGHPRGYRIOMLSFAGEPLDONS</td>	QRLQVTRKLLLEBQAAFLVGSATPRYLTLASNSHNKMGHPRGYRIOMLSFAGEPLDONS
DB <td>560 <td>QRLQVTRKLLLEBQAAFLVGSATPRYLTLASNSHNKMGHPRGYRIOMLSFAGEPLDONS</td> </td>	560 <td>QRLQVTRKLLLEBQAAFLVGSATPRYLTLASNSHNKMGHPRGYRIOMLSFAGEPLDONS</td>	QRLQVTRKLLLEBQAAFLVGSATPRYLTLASNSHNKMGHPRGYRIOMLSFAGEPLDONS
QY <td>601 <td>SMARGFSEMERVYOLAVTQRKEEPPSSSVFNONDPMAATVPSDFINNETIAGKDLVAVMT</td> </td>	601 <td>SMARGFSEMERVYOLAVTQRKEEPPSSSVFNONDPMAATVPSDFINNETIAGKDLVAVMT</td>	SMARGFSEMERVYOLAVTQRKEEPPSSSVFNONDPMAATVPSDFINNETIAGKDLVAVMT
DB <td>620 <td>SMARGFSEMERVYOLAVTQRKEEPPSSSVFNONDPMAATVPSDFINNETIAGKDLVAVMT</td> </td>	620 <td>SMARGFSEMERVYOLAVTQRKEEPPSSSVFNONDPMAATVPSDFINNETIAGKDLVAVMT</td>	SMARGFSEMERVYOLAVTQRKEEPPSSSVFNONDPMAATVPSDFINNETIAGKDLVAVMT
QY <td>661 <td>AGFLIHPIHADIPNTVTVGNGVGFLLRYNFEFDEDPSTFYADSPSYFGDDAGACEVNP</td> </td>	661 <td>AGFLIHPIHADIPNTVTVGNGVGFLLRYNFEFDEDPSTFYADSPSYFGDDAGACEVNP</td>	AGFLIHPIHADIPNTVTVGNGVGFLLRYNFEFDEDPSTFYADSPSYFGDDAGACEVNP
DB <td>680 <td>AGFLIHPIHADIPNTVTVGNGVGFLLRYNFEFDEDPSTFYADSPSYFGDDAGACEVNP</td> </td>	680 <td>AGFLIHPIHADIPNTVTVGNGVGFLLRYNFEFDEDPSTFYADSPSYFGDDAGACEVNP</td>	AGFLIHPIHADIPNTVTVGNGVGFLLRYNFEFDEDPSTFYADSPSYFGDDAGACEVNP
QY <td>721 <td>ACLPQAAAACAPDLPAFSGHGRFSH 743</td> </td>	721 <td>ACLPQAAAACAPDLPAFSGHGRFSH 743</td>	ACLPQAAAACAPDLPAFSGHGRFSH 743
DB <td>740 <td>ACLPQAAAACAPDLPAFSGHGRFSH 743</td> </td>	740 <td>ACLPQAAAACAPDLPAFSGHGRFSH 743</td>	ACLPQAAAACAPDLPAFSGHGRFSH 743

Query Match
Best Local Similarity 82.8%; Score 3343; DB 1; Length 762;
Matches 619; Conservative 44; Mismatches 75; Indels 10; Gaps 2;

QY	4	VLVYGR--GGDGE-----PSQLPHCSVSPSAQPTWHPGQSOLFADLSREELTAVM
DB <td>12 <th>VLVYGR--GGDGE-----PSQLPHCSVSPSAQPTWHPGQSOLFADLSREELTAVM</th> </td>	12 <th>VLVYGR--GGDGE-----PSQLPHCSVSPSAQPTWHPGQSOLFADLSREELTAVM</th>	VLVYGR--GGDGE-----PSQLPHCSVSPSAQPTWHPGQSOLFADLSREELTAVM

Query Match
Best Local Similarity 82.8%; Score 3343; DB 1; Length 762;
Matches 619; Conservative 44; Mismatches 75; Indels 10; Gaps 2;

QY	4	VLVYGR--GGDGE-----PSQLPHCSVSPSAQPTWHPGQSOLFADLSREELTAVM
DB <td>12 <th>VLVYGR--GGDGE-----PSQLPHCSVSPSAQPTWHPGQSOLFADLSREELTAVM</th> </td>	12 <th>VLVYGR--GGDGE-----PSQLPHCSVSPSAQPTWHPGQSOLFADLSREELTAVM</th>	VLVYGR--GGDGE-----PSQLPHCSVSPSAQPTWHPGQSOLFADLSREELTAVM

```

QY 54 RFLTORLPGVLVDAQAQAPSDNCVSVLEQLPPKAAALAHLDKSGPPPARALATVEFGR 113
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 72 SFLUQKGRPLVDAQAQAPSDNCVSVLEQLPPKAAALAHLDKSGPPPARALATVEFGR 131
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 114 QOPNVSSELVGLPHPSYKRDVTVERRHGRPLPYHRRPVLEFOEYLDIDQMFNRELPOAS 173
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 132 QOPNVTVELVGLPQPSYKRDVTVERRHGRPLPYHRRPVLEFOEYLDIDQMFNRELPOAA 191
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 174 GLHHCCFYKRRGRNLVTMTTAPRGLOSGDRATWFGLYNYSAGAFGLHNGLELVNKH 233
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 192 GVLHHCSSYKGGKGLTMTTAPRGLOSGDRATWFGLYNYSAGAFGLHNGLELVNKH 251
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 234 ALDPARTIOKVFQGRYYSLAQLEAFAGLVNVLIPDNGGSGMSLSPPPGPAP 293
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 252 ALDPADMTIQVFGQRYYSLEQLEAFAGLVNVLIPDNGGSGMSLSPPPGPAP 311
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 294 PLQFYPOGRPFSYVQSGSVASSLWTFSGLAFSGPRIFDVRFOGERLYEISLOEALAIY 353
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 312 PLQFHPQGRPFSYVQSGSVASSLWTFSGLAFSGPRIFDVRFOGERLYEISLOEALAIY 371
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 354 GGNPRAMTRRYVDGGRGKMTTPTLRGYDCPYLATYVDMHFLLESQAPRTINDARCVF 413
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 372 GGNPRAMTRRYVDGGRGKMTTPTLRGYDCPYLATYVDMHFLLESQAPRTINDARCVF 431
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 414 EONQGLPLRRHSHSLYSHYFGLAETVLVYRSMSTLLNDYDMDVTFEHPSGAIEIRFYAT 473
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 432 EONQGLPLRRHSHSLYSHYFGLAETVLVYRSMSTLLNDYDMDVTFEHPSGAIEIRFYAT 491
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 474 GYISSAFLFGATGYKGNQVSEHNLGTVHTSHAFKVDLDVAGLENNWVAEDMVEFPAAPV 533
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 492 GYISSAFLFGATGYKGNQVSEHNLGTVHTSHAFKVDLDVAGLENNWVAEDMVEFPAAPV 551
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 534 WSPHOLQRLQVNFKLEMEBOAFLVGSAPRYTLASNSNMGHGRGRIGRLMSLAG 593
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 552 WSPHOLQRLQVNFKLEMEBOAFLVGSAPRYTLASNSNMGHGRGRIGRLMSLAG 611
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 594 EPLPONSMAKGSWEERYQLAVTORKEEPPSSSVFNQNDPMWATVDFSDINNETIAGK 653
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 612 RPLPONSSTERALISMGRYQLAVTORKEEPPSSSVFNQNDPMWATVDFSDINNETIAGK 671
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 654 DLVAMTAGLHPIHADIPRTVTVNGVGFLLRPYNEFEDDPSFYADSTIFRGDODAG 713
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 672 DLVAMTAGLHPIHADIPRTVTVNGVGFLLRPYNEFEDDPSFYADSTIFRGDODAG 731
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 714 ACEVNPPLACLPQAAACAPDLPFASHGCF 741
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 732 SCEINPLACLPQAAACAPDLPFASHGCF 759
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 4
AOCX_BOVIN STANDARD; PRT; 762 AA.

```

ID AOCX_BOVIN STANDARD; PRT; 762 AA.
AC 029437;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Copper amine oxidase, liver isozyme precursor (EC 1.4.3.6) (Amine
OS oxidase [copper-containing]) (Serum amine oxidase) (SAO).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RA MEDLINE=9419366; PubMed=8144587;
RA Burilongame A.L., Smith A.J., Cai D., Kilman J.P.;
RT "Primary structures for a mammalian cellular and serum copper amine
oxidase";
RL J. Biol. Chem. 269:9926-9932(1994).

```

```

RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=9309748; PubMed=1457410;
RA Jones S.M., Palcic M.M., Scaman C.H., Smith A.J., Brown D.E.,
RA Dooley D.M., Mure M., Kilman J.P.;
RT "Identification of topaquinone and its consensus sequence in copper
RT amine oxidases";
RL Biochemistry 31:12147-12154(1992).
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
CC H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinone per subunit.
CC -1- ENZYME REGULATION: INHIBITED BY AMINOACIDS IN A COMPETITIVE MANNER.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: Liver.
CC -1- PM: Topaquinone (TPO) is generated by copper-dependent
CC autooxidation of a specific tyrosyl residue (by similarity).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S69583; AAB30397.1; -.
DR EMBL: L27218; AAB30525.1; -.
DR InterPro: IPR000269; CUNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
DR Pfam: PF02727; Cu_amine_oxid2; 1.
DR Pfam: PF02728; Cu_amine_oxid3; 1.
DR PRINTS: PR00766; CUDAOXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
KW Oxidoreductase; Copper; TPO; Glycoprotein; Signal; Metal-binding.
FT SIGNAL 1
FT CHAIN 17 762 COPPER AMINE OXIDASE, LIVER ISOZYME.
FT MOD_RES 470 470 TOPAQUINONE (BY SIMILARITY).
FT METAL 519 519 COPPER (POTENTIAL).
FT METAL 521 521 COPPER (POTENTIAL).
FT METAL 521 521 COPPER (POTENTIAL).
FT BINDING 672 683 AMINOACIDS (BY SIMILARITY).
FT BINDING 672 683 AMINOACIDS (BY SIMILARITY).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 665 665 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 762 AA; 84756 MW; AA959713602955FE CRC64;

```

Query Match 83.1%; Score 3331; DB 1; Length 762;
Best Local Similarity 81.4%; Pred. No. 7.1e-254;
Matches 609; Conservative 59; Mismatches 70; Indels 10; Gaps 2;

```

QY 4 VLVNRR--GGGGE-----PSQLPCHPSVSSAQPWTHPPGSQLFADLSREELTAVM 53
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 12 LTVNGREGDGVSEGVGKCHPPLPRCPSPSPSDPMTTPDQSLFADLSREELTAVM 71
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 54 RFLTORLPGVLVDAQAQAPSDNCVSVLEQLPPKAAALAHLDKSGPPPARALATVEFGR 113
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 72 SFLUQKGRPLVDAQAQAPSDNCVSVLEQLPPKAAALAHLDKSGPPPARALATVEFGR 131
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 114 QOPNVSSELVGLPHPSYKRDVTVERRHGRPLPYHRRPVLEFOEYLDIDQMFNRELPOAS 173
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 132 QOPNVTVELVGLPQPSYKRDVTVERRHGRPLPYHRRPVLEFOEYLDIDQMFNRELPOAA 191
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 174 GLHHCCFYKRRGRNLVTMTTAPRGLOSGDRATWFGLYNYSAGAFGLHNGLELVNKH 233
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 192 GVLHHCSSYKGGKGLTMTTAPRGLOSGDRATWFGLYNYSAGAFGLHNGLELVNKH 251
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 234 ALDPARTIOKVFQGRYYSLAQLEAFAGLVNVLIPDNGGSGMSLSPPPGPAP 293
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 252 ALDPADMTIQVFGQRYYSLEQLEAFAGLVNVLIPDNGGSGMSLSPPPGPAP 311
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 294 PLQFYPOGRPFSYVQSGSVASSLWTFSGLAFSGPRIFDVRFOGERLYEISLOEALAIY 353
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

DB 312 PLOFPPGRRSVQGRNRASSLMTFSFGLGAFSGPRVFVROGGERLAYEISLOEAGAY 371
OY 354 GGNSPAAMTTRYVYDGFSGKYYTTPTRGVCDPYLATYVDMHFLLESQAPKTRDAFCVF 413
DB 372 GGNTPAAMLTRYMDSGFGNGYFATPLIRGDCPYLATYVDMHFLVESQPKTLHDAFCVF 431
OY 414 EONOGILPLRRHSDLSXHFGLAEVLYVRSKSTILANDYVDMVYFHSGLAIEIFYAT 473
DB 432 EONKGLPLRRHSDLSXHFGLAEVLYVRSKSTILANDYVDMVYFHSGLAIEIFYAT 491
OY 474 GYISSAFELGATGKNGVSEHTLGTVTHSAHFKYDVLAVGLENNVMAEDVAFVMAVP 533
DB 492 GYISSAFELGATGKNGVSEHTLGTVTHSAHFKYDVLAVGLENNVMAEDVAFVMAVP 551
OY 534 WSPHOLOQLVYTRKLEMEQAAFLVGSATPRYLYLASNHSKMGHPRGYRIQMLSPAG 593
DB 552 WSPHOLOQLVYTRKLEMEQAAFLVGSATPRYLYLASNHSKMGHPRGYRIQMLSPAG 611
OY 594 EELPQSSMARGFSGMYGLATYQKKEEPSSSVFNQNDPAPYVDFDFITNETIACK 653
DB 612 GMPQNSPERAFSGWGYLATYQKKEEPSSSVFNQNDPAPYVDFDFITNETIACK 671
OY 654 DLVAVMTAGFLHPIPHADIPNTVYGVGFELRPYFDEDFSFYSADSIYPRGQDAG 713
DB 672 DLVAVMTAGFLHPIPHADIPNTVYGVGFELRPYFDEDFSFYSADSIYPRGQDAG 731
OY 714 ACEVNPACLPQAAACAPDLPAFSGGF 741
DB 732 SCEINPLACLPQAAACAPDLPAFSGGF 759

RESULT 5
AOC2_HUMAN
ID AOC2_HUMAN STANDARD; PRT; 756 AA.
AC 075106; 075105; 000120; Q9UNY0;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Retina-specific copper amine oxidase precursor (EC 1.4.3.6) (RAO)
GN AOC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=97237047; PubMed=9119395;
RA Imanura Y., Kubota R., Wang Y., Asakawa S., Kudoh J., Mashima Y.,
RA Oguchi Y., Shimizu N.;
RT "Human retina-specific amine oxidase (RAO): cDNA cloning, tissue
RT expression, and chromosomal mapping.";
RL Genomics 40:277-283(1997).
RN [2]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=98390194; PubMed=9722954;
RA Imanura Y., Noda S., Mashima Y., Kudoh J., Oguchi Y., Shimizu N.;
RT "Human retina-specific amine oxidase: genomic structure of the gene
RT (AOC2), alternatively spliced variant, and mRNA expression in
RT retina.";
RL Genomics 51:293-298(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RA Zhang X., McIntire W.S.;
RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY BE A CRITICAL MODULATOR OF SIGNAL TRANSMISSION IN
CC RETINA, POSSIBLY BY DEGRADING THE BIOGENIC AMINES DOPAMINE,
CC HISTAMINE, AND PUTRESCINE.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +

```

```

CC H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaguinone per subunit.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
CC HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: RETINAL-SPECIFIC.
CC -1- PTM: Topaguinone (TPQ) is generated by copper-dependent
CC autoxidation of a specific tyrosyl residue (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAGUINONE OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AB012943; BAA32590.1; -.
DR EMBL: AB012942; BAA32590.1; JOINED.
DR EMBL: AB012943; BAA32589.1; -.
DR EMBL: AB012942; BAA32589.1; JOINED.
DR EMBL: D88213; BAA19001.1; -.
DR EMBL: AF081363; AAD39345.1; -.
DR Genew: HGNC:549; AOC2.
DR MIM: 602268; -.
DR InterPro: IPR000269; CUNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid. 1.
DR Pfam: PF02727; Cu_amine_oxid2; 1.
DR Pfam: PF02728; Cu_amine_oxid3; 1.
DR PRINTS: PR00766; CUDAOXIDASE.
DR PROSITE: PS01164; COPPER AMINE OXID-1; 1.
DR PROSITE: PS01165; COPPER AMINE OXID-2; 1.
KW Oxioreductase; copper; TPQ; Alternative splicing; Glycoprotein;
FT SIGNAL 1 32
FT CHAIN 33 756
FT MOD_RES 465 465
FT METAL 516 516
FT METAL 518 518
FT METAL 680 680
FT BINDING 669 669
FT CARBOHYD 133 133
FT CARBOHYD 198 198
FT CARBOHYD 226 226
FT CARBOHYD 588 588
FT CARBOHYD 662 662
FT VARSPLIC 599 625
FT CONFLICT 181 181
FT CONFLICT 215 218
FT CONFLICT 221 222
FT CONFLICT 610 610
SQ SEQUENCE 756 AA; 83793 MW; 4E3B7317E6DAC66 CRC64;

Query Match 64.0%; Score 2565; DB 1; Length 756;
Best local Similarity 65.3%; Pred. No. 1,1e-193;
Matches 485; Conservative 88; Mismatches 162; Indels 8; Gaps 3;

OY 1 LVCVLLVGRGCGDEPSQLPCHPSVSPSAQPMWTHPGOSQLFADLSREELTAVNRELTQRL 60
DB 20 LAVVLLTSPGSGS----SQPHCPVSYSRAQPMWHPGQSGLFADLSREELTAVNRELTQRL 75
OY 61 GPGVLDAQAQRPSDNCVFSVELDLPKAAALAHLDKSGPPARADALAIYFSGQPOPQNS 120
DB 76 GPGVLDAQAQAQPSDNCIFSVELDLPKAAALAHLDKSGPPARADALAIYFSGQPOPQNS 135
OY 121 ELVVGPLPSPSYKRDVTVRHGGPLPYHRRPVYFQEYLDIDQIFNELPQASGLHHCC 180
DB 136 ELVVGPLPSPSYKRDVTVRHGGPLPYHRRPVYFQEYLDIDQIFNELPQASGLHHCC 194
OY 181 FYKRRGNLTVMTAPRGLOSGDRATWFGLYNISGAGFLVHVGLELLVNHKRALDPAW 240
DB 195 -FMYNSTLAANATPRGLSRETTMMALYHNISGVLGLHPVGLLELLDHRALDPAW 253

```

```

OY 241 TIOKFGYRYYDSLAOLEAOFAGLVNVLIPDNQGTSGSLKSPVPPAPLQFPQ 300
DB 254 TIOVQFYLYHYADLOGLEREFKSGRLVRYRPLPPNKAASSLRNSGGLPPLDQFSPQ 313
OY 301 GPRFVSQGSRAVSLMTSFGGLAFSGPRIFDVRFQGERLYEISLOALATYIGNSPPAA 360
DB 314 GSQYVQGNLVLVSLMSFFGHVFGSLIFDVRFQGERLYEISLOALATYIGNSPPAA 373
OY 361 MTRRVVDGFGGKRTTPTLRVDCPYLATYDWHFLESQAPKTRDAFCVFEONQULP 420
DB 374 MTRRLDSSFGIGRNSRGVGVDCPYQATVWDHILVKGAVOLLPGAVCVFEONQULP 433
OY 421 LRRHSDSYSHFYGSLAETVLVVRSMSTLLNDYVMDTFHPSGALIEIFRYATGYSSAF 480
DB 434 LRRHNYLQNHRYGLASALVYRVSNGYDITMDVLYNGALGEGVHATGYTNTAF 493
OY 481 LFGATG--KYGNQVSEHTLGTVHTSHAFKVDLDVAGLENNVMAEDMVFVPAVPMSPH 538
DB 494 LKSGEGGLLFGNRVGERVLTGVTHAFHFKLDLDVAGLKNVVAEDVYKPAAPMNPBH 553
OY 539 OLORLOVTRKLEMEQOAFVVGSAFPRYLYLASNHSNKGHPRGRIOMLSFAGEPLPQ 598
DB 554 WLORPOLTRQVYKEDLTFASGLSPLPRYLYLASNOTNMGHQRGRIQIHSPGLQIPL 613
OY 599 NSSMARGFSEWERYQAVYQKKEEPPSSSVFNQNDPMAFVDFSPFINNETIAGKDLVAM 658
DB 614 ESDMERALSMGXYQLVYQKKEESSOSIYHONDIWPTVTFADFINNETIAGKDLVAM 673
OY 659 VTAGLHLPHEADINVTYVNGVGFRLRPVNFEDDPSFASDSTYERGDADAGAEVY 718
DB 674 VTASLHLPHEADINVTYVNGVGFRLRPVNFEDDPSFASDSTYERGDADAGAEVY 733
OY 719 PLACLPQAAACAPDLPAFSGHF 741
DB 734 PVACLPDLAACVDPDLPPSYHGF 756

```

RESULT 6

ABP_RAT ID STANDARD: PRF: 746 AA.

```

AC P36633:
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Amiloride-sensitive amine oxidase [copper-containing] precursor
DE (EC 1.4.3.6) (Diamine oxidase) (DAO) (Amiloride-binding protein)
DE (ABP) (Histaminase).
CN ABP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Colon, and Lung;
RX MEDLINE=93387321; PubMed=8375402;
RA Linguella E., Renard S., Volley N., Waldmann R., Chassande O.,
RA Lazdunski M., Barbry P.;
RT "Molecular cloning and functional expression of different molecular
RT forms of rat amiloride-binding proteins."
RL Eur. J. Biochem. 216:679-687(1993).
CC -1- FUNCTION: CATALYZES THE DEGRADATION OF COMPOUNDS SUCH AS
CC PUTRESCINE OR HISTAMINE.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +
CC H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- PM: Topaquinine (TPO) is generated by copper-dependent
CC autooxidation of a specific tyrosyl residue (By similarity).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

```

CC EMBL: X73911; CAA52116.1;
CC DR EMBL: X73912; CAA52117.1; ALT_INT.
CC DR PIR: S34656; S34656.
CC DR PIR: S36847; S36847.
CC DR InterPro: IPR000269; CUNH_oxidase.
CC DR Pfam: PF01179; Cu_amine_oxid.
CC DR Pfam: PF02127; Cu_amine_oxidn2.
CC DR Pfam: PF02728; Cu_amine_oxidn3.
CC DR PRINTS: PR00766; CUDAOXIDASE.
CC DR PROSITE: PS01164; COPPER AMINE OXID.
CC DR PROSITE: PS01165; COPPER AMINE OXID.
CC KW Signal: Glycoprotein; Oxidoreductase; Copper; Heparin-binding; TPO;
KW Metal-binding.
FT SIGNAL 23 746
FT CHAIN 1 746
FT MOD_RES 456 456
FT METAL 505 505
FT METAL 507 507
FT METAL 670 670
FT BINDING 563 570
FT BINDING 659 659
FT CARBOHYD 110 110
FT CARBOHYD 269 269
FT CARBOHYD 533 533
FT CARBOHYD 740 740
FT SEQUENCE 746 AA; 85021 MW; 6C564D044D07BCB2 CRC64;

```

Query Match 34.9%; Score 1398.5; DB 1: Length 746;

Best local similarity 40.3%; Pred. No. 5.2e-102;

Matches 294; Conservative 127; Mismatches 266; Indels 43; Gaps 13;

```

OY 37 QGOLFADLSREELTAVMRELQRLGRLVDAQAARPSDNCVSEVLEQLPPKAAALAHDR 96
DB 29 KARVADLSPOEIKAVHSLNREELQPSKEPTLAKNSVFLIELMLPKKHVLFKDE 88
OY 97 GSPPARRELAIVFGRQOPRVSVSLVGPLRPHSYMDVYERHGGPLRYRRVYLOE 156
DB 89 GKGPRAVAVIFPGAQDYVAVTEFVAVGLPRPYIRALS-PRGHHLSWSRISIAE 147
OY 157 YLDIDOMIFNRELPOASGLIHH-----CFYKRGRLVMTAPRGLQSGDRATWFG 209
DB 148 Y----DLVYH-TLKRATMPLHGFIDTGFSLGDDDLCTTVDVAPRGVASGGRSMFI 202
OY 210 LYTNISGAGFELHVGLELVNKKALDPAKWTIOKVFYQGRYDSLAOLEAAGLVNV 269
DB 203 VQRYVE--GYFLHPTGLELLDHGSTDVODMVRVLEDMVNGKRYNPEELARKYAVGEVDT 260
OY 270 VLIIP-----NGTSGMSLSKSPV-----PGPARPDLQFYQGRFVSQGSRAVSS 314
DB 261 VVLEDPPLPVGTE-----KPLFSSTKPRGEHTPVNVAQPHVQSPGSPYKLEGTVLYG 315
OY 315 LMTFSGGLAFSGPRIFDVRFQGERLYEISLOALATYIGNSPPAAMTRYVDGFGMGK 374
DB 316 GMSFGRYLRSSGGLDIPNVLFGCEGVAEVSQAEVALYGGHTPRAGMOKTYIDVGMGLGS 375
OY 375 YTPPLTRGVDCPYLATYDWHFLESQAPKTRDAFCVFEONQULPRLRRHSDLS---H 431
DB 376 VTHELPAGIDCPETATFADFHYDSDGPRVHYPHALCFEMPTGVPLRRHFSNKGGFN 435
OY 432 YEGGAEVLYVRSMTLLNDYVMDTFHPSGALIEIFRYATGYSSAFLEGATGKYNQ 491
DB 436 FYAGLKGVLYLTSTVYNTDYIDWTFYSGVGEAKMHAGYHAFTYPRGIRHGT 495
OY 492 VSEHTLGTVHTSHAFKVDLDVAGLENNVMAEDMVFVPAVPMSPHOLQVTRKLE 551

```


DB 496 LQTHLLGNHTLVHVRVDNVAAGTAKNSFQTLTKLNTLNTPWSPSHSLVQPTLEOTQYS 555
 QY 552 MEEQAFVLSGATPRYLTLASNHNKNGHPRGYRIOMLSPAGEPLPONSSMARQSWERY 611
 DB 556 QEQHAAFRFGQGLPKYLPLSSPOKNCNGHRSTRLOIHSNAEYLVLPNGQEEARVYTRY 615
 QY 612 QLAVTORKEEPPSSSVFNONDPWAPTVDFSDFI-NNETIAGKDLVAVMTAGFLHPIHAE 670
 DB 616 PLAVTKYRESERYSSSLYNQNDPMDPVVEEFELRNMENTIETEDLVAVMTVGLFHPHSE 675
 QY 671 DINTVTYVNGVGFELRPVNFPEDEPSYASDIYFGDDAGACEVNPACLPQAACA 730
 DB 676 DVNRTATPGNSVGFLLRPNFEPEDBSLASRDVTVV-WPDCKGLNRVQ--RWIPEDRCL 732
 QY 731 PDLPAFSGG 740
 DB 733 VS-PFSSYNG 741

RESULT 7
 ABP_HUMAN
 ID ABP_HUMAN STANDARD: PRT: 751 AA.
 AC P19801: Q16683; Q16684;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Amloride-sensitive amine oxidase [copper-containing] precursor
 DE (EC 1.4.3.6) (Diamine oxidase) (DAO) (Amloride-binding protein)
 DE (ABP) (Histaminase).
 GN ABP1 OR DAO1 OR AOCI.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-94237856; PubMed-8182053;
 RA Chassande O., Renard S., Barbry P., Lazdunski M.;
 RT "The human gene for diamine oxidase, an amloride binding protein.
 RT Molecular cloning, sequencing, and characterization of the promoter.";
 RL J. Biol. Chem. 269:14484-14489(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX TISSUE-Placenta;
 RX MEDLINE-96113540; PubMed-8595053;
 RA Zhang X., Kim J., McIntire W.S.;
 RT "cDNA sequences of variant forms of human placenta diamine oxidase.";
 RL Biochem. Genet. 33:261-268(1995).
 RN [3]
 RP PRELIMINARY SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX TISSUE-Kidney;
 RX MEDLINE-91017502; PubMed-2217167;
 RA Barbry P., Champe M., Chassande O., Munemitsu S., Champigny G.,
 RA Lingueglia E., Maes P., Frelin C., Tatar A., Ullrich A.,
 RA Lazdunski M.;
 RT "Human kidney amloride-binding protein: cDNA structure and functional
 RT expression.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:7347-7351(1990).
 RN [4]
 RP SEQUENCE OF 20-39, AND CHARACTERIZATION.
 RX TISSUE-Placenta;
 RX MEDLINE-94193685; PubMed-8144586;
 RA Novotny M.F., Chassande O., Baker M., Lazdunski M., Barbry P.;
 RT "Diamine oxidase is the amloride-binding protein and is inhibited by
 RT amloride analogues.";
 RL J. Biol. Chem. 269:9921-9925(1994).
 CC -1- FUNCTION: CATALYZES THE DEGRADATION OF COMPOUNDS SUCH AS
 CC PUTRESCINE, HISTAMINE, SPERMINE, AND SPERMIDINE, SUBSTANCES
 CC INVOLVED IN ALLERGIC AND IMMUNE RESPONSES, CELL PROLIFERATION,
 CC TISSUE DIFFERENTIATION, TUMOR FORMATION, AND POSSIBLY APOPTOSIS.
 CC PLACENTAL DAO IS THOUGHT TO PLAY A ROLE IN THE REGULATION OF THE
 CC FEMALE REPRODUCTIVE FUNCTION.
 CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +

CC H(2)O(2).
 CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
 CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: PLACENTA AND KIDNEY.
 CC -1- PTM: Topaquinine (TPQ) is generated by copper-dependent
 CC autooxidation of a specific tyrosyl residue.
 CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPOQUINONE OXIDASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 DR EMBL: X78212; CA55046.1; -;
 DR EMBL: U11862; AAC50270.1; -;
 DR EMBL: U11863; AAB60381.1; -;
 DR EMBL: M55602; AAB58358.1; ALT_SEQ.
 DR PIR: A38276; A38276.
 DR PIR: S42495; S42495.
 DR Genew; HGNC:80; ABP1.
 DR MIM: 104610; -;
 DR InterPro: IPR000269; CuNH_oxidase.
 DR Pfam: PF01179; Cu_amine_oxid; 1.
 DR Pfam: PF02727; Cu_amine_oxid2; 1.
 DR Pfam: PF02728; Cu_amine_oxid3; 1.
 DR PRINTS: PR00766; CU DAOXIDASE.
 DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
 DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
 DR Signal; Glycoprotein; Oxidoreductase; Copper; Heparin-binding; TPQ;
 DR Alternative splicing; Polymorphism; Metal-binding.
 KM SIGNAL
 FT CHAIN 1 19
 FT 20 751
 FT METAL 391 391
 FT MOD_RES 461 461
 FT METAL 510 510
 FT METAL 512 512
 FT METAL 675 675
 FT BINDING 568 575
 FT BINDING 664 664
 FT CARBOHYD 110 110
 FT CARBOHYD 168 168
 FT CARBOHYD 538 538
 FT CARBOHYD 745 745
 FT VARSPLIC 619 619
 FT VARIANT 645 645
 FT CONFLICT 28 28
 FT CONFLICT 332 332
 FT SEQUENCE 751 AA; 85363 MW; 1B8B695C44232DA CRC64;
 Query Match 34.8%; Score 1394; DB 1; Length 751;
 Best Local Similarity 39.8%; Pred. No. 1.2e-101;
 Matches 293; Conservative 124; Mismatches 261; Indels 38; Gaps 11;
 QY 31 PWTHPGOSQFADYSREELTVMRFLQRLPGVLDAQAARPSDNCVSVLELQPPKAA 90
 DB 23 PGLTPRKAAGVSDSNOLKAVHFLWMSKELRLQPSSTTMANKTVFLIEMLPKKYHV 82
 QY 91 LAHLDGSPPARALAVFEGRQPPVSVSLVGLPHRYMYRVTERRHGGPLPYRR 150
 DB 83 LRFIDGRHPRVREARAVIFGDDEHPVTEFAVGLPGCYMRALS-PRGYOSSWASR 141
 QY 151 PVLQOEYLDIDQMFNRELPOASGLLH-----CCFYKHGRRLVTMTAPRGLQSGD 203
 DB 142 PISTAEY-----ALLXHLQELATKPLHOFPLNTGFSQDCHDLATVDVAPRGVASQ 196
 QY 204 RATWFGLYYNSIGAGFFLHVHGLLELVNKKALDPAKMTIQKVFYGRYDSLAOLEA 263

```
Db 197 RRSMLIIQRYVE--GYFLHPTGLLELVHDSGTADAGHMAVEQWYNGKKEFGSPHEELARKYA 254
QY 264 AGLVNVLIIIPNGCGSMKSLSPVPR-----GAP-----PLOGYPCGPPFPSSVQ 308
Db 255 DGEVDVNVLEPRLPGCKHSDSTEEPPLESSHKPRGDFSPSPHVSGRVLVQRPFRFRLEG 314
QY 309 SRVASSLTFFSGGAFSGPRIFDVRFOGGERLYVEISLOEALATIGNSPAMTTRRYVDG 368
Db 315 NAVLYGSGSFAPRLRSSGLOLVANHFGEGERATVEVSQVQENALVGGHTRPAGOMOKYLDV 374
QY 369 GFGMKRYTPTLRGVDCPYLATVYDMHFLDSQAPRTIRDAFCVEQONGLPLRRHSD 428
Db 375 GWSGLSVYHEHAPGIDCEPATFTDTEFYHVDADDEVHYPRALCLPEMPTGVPLRRHNSN 434
QY 429 YS---HYEGGLAEPLVYRSMSTLLNDYVMDTVPHPSGAIEIRYANGYISSALFGAT 485
Db 435 FKGGFNFAAGKGOVLVATRTSTYVKNYDIWDFETYPNGVWEAKKAHAGYVATYTPREG 494
QY 486 GKYNQVSEHTLGTIVHTSHAFKVDLDVAGLENNWMAEDMVFVPAVMPSPERHQLRLQV 545
Db 495 LRHGTRLHTLHIGNHTLHVHYRVDLDVAGTKNSFQTLQMKENTNTNPMSPRHRYVQPTL 554
QY 546 TRKLEMEQAAFLVGSATPRILYLAHNSKMGHPRGRTIDMSLFACEPLPONSMAWG 605
Db 555 EOTOYSWRQAAAFRRKRLPYLLFTSPQENPMGHRKSRVRLQIHSMAQOVLPFGMOEQA 614
QY 606 FSWERYOLAVTORKEEPPSSSVFNQNDPMAPTVDSDFI--NNETIAGDIAVMTAGTL 664
Db 615 ITWARYPLAVTKYRRESELCSSITIHQNDPMHPVVEEDGLHNNENIEDLVAWTVGTL 674
QY 665 HIPHAEIDPNTVYNGVGFELRPYFDEDESPYSADSIYFGDQDAGACEVNPACLP 724
Db 675 HIPHSEDIPNATPCNSVGFLLRPENFPEDEPSLASRDTVLY-WRDNGPNVYQ--RWLP 731
QY 725 QAACAPDLPAFSHOG 740
Db 732 EDROCSMP-PPFSYNG 746

RESULT 8
AOC3_RAT STANDARD: PRT: 321 AA.
ID AOC3_RAT 008590;
AC 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Membrane copper amine oxidase (EC 1.4.3.6) (Vascular adhesion protein-1) (VAP-1) (Vp97) (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-Sprague-Dawley; Tissue-Adipocyte;
RX MEDLINE=97238878; PubMed=9083076;
RA Morris N.J., Ducet A., Aebersold R., Ross S.A., Keller S.R., Lienhard G.E.;
RT "Membrane amine oxidase cloning and identification as a major protein in the adipocyte plasma membrane."
RL J. Biol. Chem. 272:9388-9392(1997).
RN (2)
RP SEQUENCE OF 1-19.
RX TISSUE-Adipocyte;
RX MEDLINE=96010417; PubMed=8520629;
RA Jochen A., Given S., Hays J.;
RT "The major integral membrane glycoprotein in adipocytes is a novel 200-kDa heterodimer."
RL Mol. Membr. Biol. 12:277-281(1995).
CC -1- FUNCTION: CELL ADHESION PROTEIN THAT PARTICIPATE IN LYMPHOCYTE RECIRCULATION BY MEDIATING THE BINDING OF LYMPHOCYTES TO PERIPHERAL LYMPH NODE VASCULAR ENDOTHELIAL CELLS IN AN L-SELECTIN-
```

```
CC INDEPENDENT FASHION. HAS A MONAMINE OXIDASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) + H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
CC -1- PM: Topaquinine (TPQ) is generated by copper-dependent autooxidation of a specific tyrosyl residue (by similarity).
CC -1- PM: N- AND O-GLYCOSYLATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or\_send\_an\_email\_to\_license@sib-sib.ch).
CC -----
DR EMBL: U72632; AAC53189.1; -.
DR InterPro: IPR000269; CuNH_oxidase.
DR Pfam: PF02722; Cu_amine_oxidn; 1.
DR Pfam: PF02728; Cu_amine_oxidn; 1.
DR PRINTS: PR00766; CUDAOXIDASE.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; PARTIAL.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; PARTIAL.
KW Oxidoreductase; Copper; TPQ; Glycoprotein; Transmembrane;
KW Signal-anchor; Cell adhesion; Metal-binding.
FT INIT_MET 0
FT DOMAIN 1 5
FT TRANSMEM 6 26
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
FT DOMAIN 27 >321
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 321 321
SQ SEQUENCE 321 AA; 35116 MM; 6B1F294253A1DE6F CRC64;

Query Match 33.8%; Score 1354; DB 1; Length 321;
Best Local Similarity 83.2%; Pred. NO. 5e-99;
Matches 252; Conservative 18; Mismatches 33; Indels 0; Gaps 0;

QY 1 LVCVLVGRGSGDGEPSOLPHCPSPVSPSAOPWTHGQSOLFADLSREELTAVMRELTQRL 60
DB 19 LVCVLVLAGRSGGGRSLQPLHCPSPVSPVQPTHSGGQGFADLSPEELTAVMSLKLHL 78
QY 61 GPGLYDAQAARPSDNCVSVELQLPKAAALAHLDKSGPPAREKALAVFFGRQPPVNS 120
DB 79 GGLVDAQAARPSDNCVSVELQLPKAAALAHLDKSGPPAREKALAVFFGRQPPVNS 138
QY 121 ELVVGPLPHPSYMRVYTERHRCGPLPYHRRPYLPQEVYLDIDIMFNRELPOASGLHHCC 180
DB 139 ELVVGPLPHPSYMRVYTERHRCGPLPYHRRPYLPQEVYLDIDIMFNRELPOASGLHHCC 198
QY 181 FYKHGRNLYVTTPARGLQSGDRATWGLYYNISGAFLEHHVGLLELVNKAALDPARW 240
DB 199 FYKHGRNLYVTTPARGLQSGDRATWGLYYNISGAFLEHHVGLLELVNKAALDPARW 258
QY 241 TIQKVFYGGRIYDLSLAQLEAOFAGLVNVLIPDNGTSGMSKSPVPPGAPPLQGYTPQ 300
DB 259 TIQKVFYGGRIYDLSLAQLEAOFAGLVNVLIPDNGTSGMSKSPVPPGAPPLQGYTPQ 318
QY 301 GPR 303
DB 319 GPR 321

RESULT 9
AMO_KLEAE STANDARD: PRT: 755 AA.
ID AMO_KLEAE P49250;
AC P49250;
DT 01-FEB-1996 (Rel. 33, Created)
```

01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Copper amine oxidase precursor (EC 1.4.3.6) (Monamine oxidase)
 DE (Tyramine oxidase).
 GN MAOA OR TYNA.
 OS Klebsiella aerogenes.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Klebsiella.
 CC NCBI_TaxID=28451;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-41.
 RC STRAIN=M70;
 RX MEDLINE=92210491; PubMed=1556068;
 RA Sugino H., Sasaki M., Azakami H., Yamashita M., Murooka Y.;
 RT "A monamine-regulated Klebsiella aerogenes operon containing the
 RT monamine oxidase structural gene (maoA) and the maoc gene.";
 RL J. Bacteriol. 174:2485-2492(1992).
 CC -1- FUNCTION: ACTIVE ON TYRAMINE, TRYPTAMINE, BETA-PHENETHYLAMINE AND
 CC DOPAMINE.
 CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
 CC H(2)O(2).
 CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- INDUCTION: BY TYRAMINE AND CATECHOLAMINES.
 CC -1- PTM: Topaquinine (TPO) is generated by copper-dependent
 CC autooxidation of a specific tyrosyl residue (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUINONE OXIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: D10208; BAA01060.1; .
 DR HSSP: P46883; 10AC.
 DR InterPro: IPR000269; CUNH_Oxidase.
 DR Pfam: PF01179; Cu_amine_oxid; 1.
 DR Pfam: PF02727; Cu_amine_oxidn2; 1.
 DR Pfam: PF02728; Cu_amine_oxidn3; 1.
 DR PROSITE: PS01164; COPPER AMINE_OXID_1; 1.
 DR PROSITE: PS01165; COPPER AMINE_OXID_2; 1.
 KW Oxidoreductase; Copper; TPO; Periplasmic; Signal; Metal-binding.
 FT SIGNAL 1 30
 FT CHAIN 1 755 COPPER AMINE OXIDASE.
 FT ACT_SITE 413 413 CATALYTIC BASE (BY SIMILARITY).
 FT MOD_RES 496 496 TOPAQUINONE (BY SIMILARITY).
 FT METAL 554 554 COPPER (POTENTIAL).
 FT METAL 556 556 COPPER (POTENTIAL).
 FT METAL 719 719 COPPER (POTENTIAL).
 FT METAL 719 719 COPPER (POTENTIAL).
 SQ SEQUENCE 755 AA; 83576 MW; 785552283CD93EFF CRC64;
 Query Match 10.3%; Score 413; DB 1; Length 755;
 Best Local Similarity 24.1%; Pred. No. 1.5e-24;
 Matches 188; Conservative 101; Mismatches 284; Indels 206; Gaps 32;

234 ALDPARTYQK---FYGRYDLSLAQLAAGEAGLVNVLIPDNKGGSSW----- 282
 DB 244 ITDPGKVVVTPPLVGFEDK--DGLQQ-----DARLLKVVSYLDPTDGGYMAPIENLVA 296
 QY 283 ---LKS-----VPPGPAP-----PLOG-YPGPRFSVSGSRVAS 314
 DB 297 VVDLEAKKIIKIEEGVIVPMEPRPDGRNAPVAKPLEITEPEGKNTITGDTIHQ 356
 QY 315 LWTFSGLGAFSGPRIFDYRFQS---ERLVEISLQALATYGNSPAAMTRYVDG- 369
 DB 357 NNDPFLRLSRGPIILSTYNDNGKRGVMYEGSLGMIIVPGDDVGMFKAYIDSGD 416
 QY 370 FGKGYTPTLTGVDPCPYATYVDWHFLLESQAQKTRDAFCVFEONGPLPRHSDLY 429
 DB 417 YCMGTLTSPVKGKAPSNVLLDETADYTGKPTTPGVAIFE-----RYAGPEY 468
 QY 430 SHYFGS---LAETLVYRSMSTLLNVDYMDTVPHSGAIEIRFATGYISSAFLFGA 484
 DB 469 KHEMGKPNVSTERRELVRWISTVGNYDIFDMVHDNGTIGDAGATGEIVKGLAK 528
 QY 485 T-----GKYNQVSEHTLTGVTHSAHEKVDLDVAGLEWVAEDMVEVPAVPS 535
 DB 529 TMHDSAKEDTRVGLIDHNIVGTHQHLNFRDLVDGENTLVAMDEPKRPT---A 585
 QY 536 PEHQLORLOVTRKLEMEQAAFLVGSATPRVLYLASNHS--NKMGHPGRIOMLSENG 593
 DB 586 GSPRTSTMOVNYTTIDSEOKAAQKEDPGTIR---LLSTNSKERNMNPVSY-QIIPYAG 640
 QY 594 --EPLPONSMA-----RGFSWERYQLAVTORKEE-----PSSSVFNQNDWAPT 638
 DB 641 GHPAATGAKFAPDEWIRYRLSFMOKLWTRHPIERPEGYPRSHDGLGYAK- 699
 QY 639 VDFSEFINNETIAGKDLVAVMTAGFLIPHAEDIPRTVTVGNGVGFRLPYNFFDDPS 697
 DB 700 -----DDESLTNHDDVWMTTGTTHVARAEEMPIMPIT-EVALALLKPMNFFDEPT 749
 RESULT 10
 AMO_ECOLI STANDARD; PRT; 757 AA.
 ID AMO_ECOLI P46883; P78153; 053008;
 AC P46883; P78153; 053008;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Copper amine oxidase precursor (EC 1.4.3.6) (Tyramine oxidase)
 DE (2-phenylethylamine oxidase).
 GN TYNA OR MAOA OR B1386.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 CC NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RA Azakami H., Yamashita M., Roh J.-H., Suzuki H., Kumagai H.,
 RA Murooka Y.;
 RT "Nucleotide sequence of the gene for monamine oxidase (maoA) from
 RT Escherichia coli.";
 RL J. Ferment. Bioeng. 77:315-319(1994).
 RN [2]
 RP SEQUENCE FROM N.A., AND X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RC STRAIN=K12 / K10;
 RX MEDLINE=9616434; PubMed=8591028;
 RA Parsons M.R., Convery M.A., Wilmot C.M., Yadav K.D.S., Blakeley V.,
 RA Corner A.S., Phillips S.E.V., McPherson M.J., Knowles P.F.;
 RT "Crystal structure of a quinoxaline: copper amine oxidase of
 RT Escherichia coli at 2-A resolution.";
 RL Structure 3:1171-1184(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;

FT CONFLICT 290 290 P -> I (IN REF. 1).
 FT CONFLICT 456 456 A -> P (IN REF. 1).
 FT CONFLICT 659 659 H -> D (IN REF. 1).
 SO SEQUENCE 757 AA; 84378 MM; 65600BCED35243DB CRC64;

Query Match 9.98; Score 397.5; DB 1; Length 757;
 Best Local Similarity 27.38; Pred. No. 2.5e-23;
 Matches 142; Conservative 74; Mismatches 211; Indels 89; Gaps 20;

246 FYQGRHYDLSLAQLEAQAFLVNVLPDNGTSGSM-----LKS 285
 259 YFDGK--DGLNQ-----DARLKAVSYLDVGGNGYMAHRIENVAVNDEQKKIYKIEG 311
 266 PVPPEP-----AP--PLQFY--PGGPRFSVOGSRVASSLMTFSFGLAGSGPR 329
 312 PVVPVPMARPEPDGDRAPAPKAPKQIIEPEGNKNTITGDMIHMMNMFHLSMNSRVGPM 371
 330 IFDVAFQG-----ERLYVEISLQELALAYGNSPAMATRYVNGG--FGMCKYTTPLTRGVD 384
 372 ISTVYINDNGTKRKVMYEGSLGGMVTPYGDPDIGMYFKAYLDSDYGMGLTSPILAKGD 431
 385 CPYLATYVDMHFLDSQAQKTRDAFCVEEQNGGLPLRRHSDLYSHYFGG-----LAET 439
 432 APSNVNLNETIADYTGVPMEIPRAIVFE-----RYAGPEKKHQMGQPNVSTER 483
 440 VLUVASMTLLNTDYVMDTVFHPSCAIEIRYATGY-----ISSAFLEGATGK---YGN 490
 484 ELVVMISTVGVNDYIFPMIFEHENGTIGDAGATGIEAVKGAKTMDHETAKDTRYG 543
 491 QVSEHTLGTVHSHAFVYDVLVAGLENNVMAEDMVPMANPMSPEHQLQLOYTRKLL 550
 544 LIDHIVTGTTHOHTNFRDLDDVGEENSLVAMPVAKPNT---AGPRTSTMVQNOINI 600
 551 EMEQQAFLVGSATPRRYLYLASNSHKNQHPGRYIOMLSFAG--EPLFONSSMA----- 603
 601 GNEQQAQKFPDGTIR--LLSNKNENRBMGNPVSY--QIIPYAGGHVPAKGAQAPADEMI 657
 604 -RGFEMERYQLAVTORKEEPESSSVF--NQNDPAPATYDFSPFINNETIAGDLYAVATA 661
 658 YHRLSEFMKQLWVTRYHGEERPEPEKYPNRSYHDTGLQGYSK--DNESLNDTDAVVMWT 715
 662 GELHPIHAEDIPNTVTVGVNGVGFRLRPYFEDEPS 697
 716 GTTHVARAEEMPIMT--EMVHTLAKPMFDEPTF 749

RESULT 11
 AMOH_ARTGO STANDARD: PRT: 683 AA.
 AC 059118;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Histamine oxidase (EC 1.4.3.6) (Copper amine oxidase).
 OS Archaeobacter globiformis.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Micrococcales; Micrococcaceae; Archaeobacter.
 NC NCB1_TaxID=1665;
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RC STRAIN-IFO 12137 / ATCC 8010;
 RX MEDLINE=95181469; PubMed=7876243;
 RA Chou Y.-H., Matsuzaki R., Fukui T., Shimizu E., Yorifuji T., Sato H.,
 Ozaki Y., Tanizawa K.;
 RT "Copper/topa quinone-containing histamine oxidase from Archaeobacter
 globiformis. Molecular cloning and sequencing, overproduction of
 precursor enzyme, and generation of topa quinone cofactor.";
 RL J. Biol. Chem. 270:4712-4720(1995).
 CC -FUNCTION: OXIDIZE HISTAMINE. OTHER AMINES INCLUDING
 PHENETHYLAMINE, TYRAMINE, TRYPTAMINE, PUTRESCINE, AND BENZYLAMINE
 ALSO SERVE AS SUBSTRATE.
 CC -CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +
 H(2)O(2).

CC -COFACTOR: Binds 1 copper ion and 1 topaquinone per subunit.
 CC -SUBCELLULAR LOCATION: Cytoplasmic.
 CC -INDUCTION: BY HISTAMINE.
 CC -PM: Topaquinone (TPQ) is generated by copper-dependent
 CC autooxidation of a specific tyrosyl residue.
 CC -SIMILARITY: BELONGS TO THE COPPER/TOPOQUINONE OXIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D38508; BAA07517.1; -;
 DR HSSP: P46881; 1A4A.
 DR InterPro: IPR000269; CuNH_oxidase.
 DR Pfam: PF01179; Cu_amine_oxid; 1.
 DR Pfam: PF02727; Cu_amine_oxid2; 1.
 DR Pfam: PF02728; Cu_amine_oxid3; 1.
 DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
 DR PROSITE: PS01165; COPPER_AMINE_OXID_2; FALSE_NEG.
 KW Oxidoreductase; Copper; TPQ; Metal-binding.
 FT INIT_MET 0
 FT MOD_RES 401 401 TOPOQUINONE.
 FT METAL 450 450 COPPER (POTENTIAL).
 FT METAL 452 452 COPPER (POTENTIAL).
 FT METAL 611 611 COPPER (POTENTIAL).
 SO SEQUENCE 683 AA; 74979 MM; 233C46C017201AAO CRC64;

Query Match 9.4%; Score 377; DB 1; Length 683;
 Best Local Similarity 24.1%; Pred. No. 8.8e-22;
 Matches 182; Conservative 107; Mismatches 343; Indels 124; Gaps 28;

20 PHCPSVPSAOPWT--HPQSQSLFADLSREELTAVMRELTQSLGSLVAAQARSQNCV 77
 6 PSTPLVQDPVPVATLVHAAQHPLEQLSAEELHEARRILAE--AGLV----- 51
 78 FSEVQLPPEKAAALHLDRGSP-----PAREALATVEFGROP-----PNVSE 121
 52 -----ESTRPAVLGLIEPPKTRTQGDVGAARLVAMLMDAQNSDLVRLSLANG 102
 122 LVVGPLPHPSYKRDVTVEHNGCPPLPHRRPVLFQETLDDQIMFNRELFOASGLLHCCF 181
 103 LVVD-----RRELNPEDG-----QLPVLEEEGIIEDIL--SEDOVMAAL----- 142
 182 YKHGRNLVTMTAP-----RGLQSGDRATWFGLYYINISGAFPLHVGELLYVNHKA 234
 143 -TARGITPAQVAVAPLSAGVFEYGNNEGKRL-----LRGLGFQDHP-----ADHW 188
 235 LDPARWTQKVEFYQGRHYDLSLAQLEAQAFLVNVLPD--NGTGSWSLKSVPGPAP 293
 189 AHPIDGLVAFVDEVNRNVHL-----IDGPPV--PEVGNVYDPAIRBELRLDLP 239
 294 PLQFYPOGPRFSVOGSRVASSLMTFSFGLAGSGPRIFDVR--FOGER--LYEISLQEA 349
 240 IEIMQPEGSFYLEGNHLSWAGMDLVGFDAEGELVHQLHSHKRRRPVHIRASISEM 299
 350 LAIVGNSPAMATTRYVNGG--FGMGKYYTPPLRGVDCPPLATYVDMHFLDSQAQKTRID 408
 300 VVPYDPSPTYSKWMONTFDSGEYLVGRDANSLKCCDCLDITYMSVVAADGPNPTTN 359
 409 AFCVEQNGGLPLRRHSDLYSHYFGLAETVLVVSMSSTLLNTDYVMDTVFHPSCAIEI 468
 360 GICHEEDAGILWK--HTDEWAGSDEVRRNRRLVYSFTTVGNYDYGFWLYLIDGTIEF 417
 468 RRYANGYISSAFLFGATGTYGNQVSEHTLGTVHTSAHFKYDLDVAGLENNVMAEDMVY 528
 418 EAKAGYVETALPDKDYAVASEIAPGLGAPYHOHLFSARLMDMDIGDANRVEELDVRL 477
 529 PAAVWMSPEHQLQLOYTRKLLMEQQAFLVGSATPRRYLYLASNS--KKMGHPGRYRQ 587

```

Db 478 PKG-PGNHG--NAFTQRTLLARESEAVRDADGAKGRVWHISNPDLSLHGHGPGYTLY 534
QY 588 MLRAGEP---LPONSSMARGFSEWERYOLAYORKKEEPPSSSVFNQNDPMAPYVDSDF 644
Db 535 P---EGNPTTLAMADDSSTASRAAFARHHLWYTRHAEELLYAGDSVNOHPCGAVLP-AYV 590
QY 645 INNETHAGKDLVAVYTAGFLIHPHAEIDPNTVYVNGVGFPLRPVNFEDDESPFSADSI 704
Db 591 AODDIDGODLVVHSHFELTHFPREDMP--IMPYDTTGFTLKPFGFDENFTLVNPSSA 648
QY 705 YFRGDODAGACEVNPPLACPOAACAPDLPAFSSHOG 740
Db 649 -----AGHCCTG-----SERHNAAPGTAVGHSG 672

RESULT 12
AMOI_ARTSI STANDARD; PRT: 648 AA.
AC 007121;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Copper amine oxidase precursor (EC 1.4.3.6) (MAOXI).
GN MAOI.
OS Arthrobacter sp. (strain P1).
OC Bacteria; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcales; Micrococcaceae; Arthrobacter.
OX NCBI_TaxID=47915;
RN [1]
RP MEDLINE=93374858; PubMed=8366046;
RA Zhang X., Fuller J.H., McIntire W.S.;
RT "Cloning, sequencing, expression, and regulation of the structural
RT gene for the copper/topa quinone-containing methylamine oxidase from
RT Arthrobacter strain P1, a Gram-positive facultative methylotroph.";
RL J. Bacteriol. 175:5617-5627(1993).
CC -1- FUNCTION: THE EXACT FUNCTION OF MAOXI IS NOT KNOWN.
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
CC H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinone per subunit.
CC -1- PFM: Topaquinone (TPQ) is generated by copper-dependent
CC autooxidation of a specific tyrosyl residue (By similarity).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPOQUINONE OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; L12983; AAA22076.1; -.
DR HSSP; P46881; IAV4.
DR InterPro; IPR000269; CUNH_oxidase.
DR Pfam; PF01179; Cu_amine_oxid; 1.
DR Pfam; PF02727; Cu_amine_oxid2; 1.
DR Pfam; PF02728; Cu_amine_oxid3; 1.
DR PROSITE; PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE; PS01165; COPPER_AMINE_OXID_2; 1.
DR OXidoreductase; Copper; TPQ; Metal-binding.
KW PROPEP
FT CHAIN 1 648 COPPER AMINE OXIDASE.
FT MOD_RES 10 385 TOPAQUINONE (BY SIMILARITY).
FT METAL 436 436 COPPER (POTENTIAL).
FT METAL 438 438 COPPER (POTENTIAL).
FT METAL 595 595 COPPER (POTENTIAL).
SQ SEQUENCE 648 AA; 72760 MW; B2F9E267492253B5 CRC64;

Query Match 9.18; Score 363; DB 1; Length 648;
Best Local Similarity 24.08; Pred. NO. 1e-20;
Matches 177; Conservative 96; Mismatches 301; Indels 164; Gaps 31;

```

```

QY 36 GQSQFLADLSREELTAVMRFLLTORGLVDVAQAQRPDSNCFVSELDLPPKAAALALHD 95
Db 12 GVSHPDLPLSRELTARAAVILKE--GPAAESFR-----FISLEKREPSKDL----- 57
QY 96 RGSPPAREALATVEFGQPOPNVSELYVGPLPHPSYRDVYVERHGSPLYHRHPVLFQ 155
Db 58 RAGVAVAEADAVLV--DRAQRSEAVVD-----LEAGTVDSKMLLAENIQPFPLD 108
QY 156 EYLDIDQIFNRELPOASGLLHCCFYKRGKRNLTMTTAPRGLOSQDRA---TWFGLYY 212
Db 109 EFACEDAC--RKDE-----VIALAKRGLNLNLDVCFEPPSVGYE 148
QY 213 NISGAGFLVHVGLELVVNHKALD-PARKTIQK--VFYQGRYDSLAOLEAFEGALVNY 269
Db 149 GEDNEGRL--WRALVFVRDADSPYAHPIENFVFX-----DLNAG--KV 191
QY 270 VLIPIVNGTGSMSLSKSPPGPAPPLQFY-----PQGRFSVQSGRVASS 314
Db 192 VRLIEDQA-----IPPSARGNTLPYVGEARLDKPLNITQDEGASFYTGHNHTWA 244
QY 315 LWTFSFGLGAFSGPRIFDFVRQGE---RLYEISLQALAIYGSNSPA-AMTTRVYDGG 369
Db 245 DMSFVGFPTREGVLVHLQKFKDQGVDRPVINRSLSEWVVPYGDTPAVQAKKNAFDSGE 304
QY 370 FQMGKVTTPLRGVDCPYLATYVDMHFLLESQAPKTIADACVFEQONGPLRRHSDLY 429
Db 305 YNIGMANSLLTGCCDCELEIRYFDGHSVDSHGPNPTINAIQCMHEEDSI-----LW 356
QY 430 SH--YFGGLAET---VLVRSMSSTLVNDYVMDVFPSPGAIEIRFATGYISSA-PLF 482
Db 357 KHPDRREGTAETRRSRKVLISFIATVAVYEAFAFYHLELDSIEFLVATGTLISAGLP 416
QY 483 GATGRYGNQVSEHRL-GTVTHSHAHFKYVDLVAGLENNVMAEDVYFVPAVPMWSEHQL- 540
Db 417 GEKNPYGOSLNNDGGLYAAIHQHMFMVVRADFELDGKNAVYEVDMEY-----PENHPT 468
QY 541 -QRQVTRKLEMEBOAFLVGSAPRRLYLXASNS--KKMGHPRXRYTOMLSFAGEPLPQ 598
Db 469 GTAFNAVDRLLLETBOKAIRKTKNEAKHREKVIANNHSHKLVNEPAVRL-----IPT 519
QY 599 NSSMARGFSEWERYOLAYORKKEEPPSSSVFNQNDPMAPYVD-----FSPFINNET--- 649
Db 520 NG-----IQDLA--RDAYVSKRAQFARNNLMTAVDRTERFAAGEPQNTGAD 567
QY 650 -----IAGDLVAVYTAGFLIHPHAEIDPNTVYVNGVGFPLRPVNFEDDEPS 697
Db 568 DGLHMTQKDRNIYDTDLVWYVTFGMHHVRLDMP--VMPRONIGFMLEPHGFENQPT 625
QY 698 FYSADSIYFRGDQDAGAC 715
Db 626 LNLPTSTTGTGEADTC 643

RESULT 13
AMOI_ARTSI STANDARD; PRT: 648 AA.
AC 007123;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Copper methylamine oxidase precursor (EC 1.4.3.6) (MAOXII).
GN MAOI.
OS Arthrobacter sp. (strain P1).
OC Bacteria; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcales; Micrococcaceae; Arthrobacter.
OX NCBI_TaxID=47915;
RN [1]
RP MEDLINE=93374858; PubMed=8366046;
RA Zhang X., Fuller J.H., McIntire W.S.;
RT "Cloning, sequencing, expression, and regulation of the structural
RT gene for the copper/topa quinone-containing methylamine oxidase from
RT Arthrobacter strain P1, a Gram-positive facultative methylotroph.";

```



```

RL J. Bacteriol. 175:5617-5627(1993).
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
CC H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
CC -1- SUBUNIT: HOMODIMER.
CC -1- INDUCTION: BY METHYLAMINE.
CC -1- PTM: Topaquinine (TPO) is generated by copper-dependent
CC autooxidation of a specific tyrosyl residue (by similarity).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUNINE OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L13990; AAA22074.1; -.
DR HSPSP; P46881; 1A4.
DR InterPro: IPR000269; CUNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
DR Pfam: PF02727; Cu_amine_oxid2; 1.
DR Pfam: PF02728; Cu_amine_oxid3; 1.
DR PROSITE: PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE: PS01165; COPPER_AMINE_OXID_2; 1.
DR OXidoreductase; Copper; TPO; Metal-binding.
KW PROPEP 1
FT CHAIN 10 648 COPPER METHYLAMINE OXIDASE.
FT MOD_RES 385 385 TOPAQUNINE (BY SIMILARITY).
FT METAL 436 436 COPPER (POTENTIAL).
FT METAL 438 438 COPPER (POTENTIAL).
FT METAL 595 595 COPPER (POTENTIAL).
SQ SEQUENCE 648 AA; 72805 MW; B2FB2787492253B5 CRC64;

Query Match 9.18; Score 363; DB 1; Length 648;
Best Local Similarity 24.08; Pred. No. 1e-20;
Matches 177; Conservative 96; Mismatches 301; Indels 164; Gaps 31.

OY 36 GOSOLFADLSREELTAVNRFLTORLPGLVDAQAQRPSCNCEVSVELQPPRAAALAHLD 95
DB 12 GVSHPLDLSTHVEIRARVAALKE--GPAASESFR-----FISYLEKPSKDDL----- 57
OY 96 RGSPPAREALAIYFFGQOPQNSVSELVVGRLPHPSYMRDLYVEHNGRPLRHRPVLQ 155
DB 58 RAGVAVAEADAVLV--DRAQARSEAVD-----LEAGVDSWKLAENIOPPFMD 108
OY 156 EYLDIDQMFNRELPOASGLLHHCFFYKRGHNLVTMTTAPGLQSGDRA--TWFGLY 212
DB 109 EFAEEDAC--RKQPE-----VIALAKGLNLMDLVCEPFSVGF 148
OY 213 NISGAGFLHVGLELLVNHKALD--PARMTQK--VFYQGRYDLSLAOLEAFEGALVNV 269
DB 149 GEDNGGRRL--MRALVYFVDEADDSRYAHPIENFIYF-----DLNAG--KV 191
OY 270 VLIPIINGGGSLSKSPYPPGAPPLQFY-----PQGRFVSQGSRVASS 314
DB 192 VRLLEDQA-----IPVPSARGNLYPYVGEARTDLPLNTQEGSAFTYTGHNVTWA 244
OY 315 LMTFFGGAFCGSPRIFDFRQGE-----RLVYEISLQELATLYGGSNSP--AMTRRVDDG 369
DB 245 DMSFVGTTPREGVLVHLQKFKDQGVDRVYNRAISLEENVYVGTAPYQAQAKNAFDSGE 304
OY 370 FGMGYTTPPLRFGVDCPYLATVYDMHFLLESQAPRTINDAFVFEQONGLPLRRHSDLY 429
DB 305 YNIGMANSALTGLGCDLGEIKFYFGDGHVSYDSCNPMTIENAIOMHEBDSI-----LM 356
OY 430 SH--YFGGLAET-----VLVYRSMSTLNTDYMDVVFHSGAIEIRFYATGYSN--FLF 482
DB 357 KHFDREGAETRRSRKLVLSIFATVAVNTEVAFYVHMLFDGSIIEFLVKATGILSTAGQLP 416
OY 483 GATGKGNQVSEHTL--GVHTSHAFKVDLVDAGLENNWADNVFVPAVWSPENHQL- 540

```

```

DB 417 GERNPYGOSLNNDGLIAPHIHOHMFVNRADFELDGKYNAYVEVDY-----PENHPT 468
OY 541 -QRLQVTRKLLMEEQAAFLVGSATPRYLALASNS--NRKGHPRCYRIOMLSFAGEPPLQ 598
DB 469 GTAFVADVRLLETQEKAIKRTNKAKHFMKLIANHSKMLVNEPVAURL-----IPT 519
OY 599 NSSMARGSWERYOLAVYORKEEPPSSSVYNDNPAPYVD-----FSDFINET--- 649
DB 520 NG-----IDLA--RQDAYVSKRAQFARNLWLTAYDRTERFAAGEYPNQATAD 567
OY 650 -----IAGKDLVAVWTAGFLIHPAEDIPNTVYGVNGVFLRPYNFDEDDPS 697
DB 568 DGLHNTQKNDINVDLTDLVWYTFEGMHVYRLDMP--VMPRONIGFMLEPHGFENQNP 625
OY 698 FYSADSIYFRGDQAGAC 715
DB 626 INLPSTSTGTQTEADTC 643

RESULT 14
PAOX_ARTGO STANDARD: PRT; 638 AA.
ID PAOX_ARTGO
AC P46881;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phenylethylamine oxidase precursor (EC 1.4.3.6) (amine oxidase).
OS Arthrobacter globiformis.
CC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
CC Actinomycetales; Micrococcales; Micrococcaceae; Arthrobacter.
OX NCBI_TaxID=1665;
[1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP STRAIN=IFO 12137 / ATCC 8010;
RC MEDLINE=94197690; PubMed=8147851;
RA Tanizawa K., Matsuzaki R., Shimizu E., Yorifuji T., Fukui T.;
RT "Cloning and sequencing of phenylethylamine oxidase from Arthrobacter
RT globiformis and implication of Tyr-382 as the precursor to its
RT covalently bound quinine cofactor."
RL Biochem. Biophys. Res. Commun. 199:1096-1102(1994).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 9-628.
RX MEDLINE=98070236; PubMed=9405045;
RA Milcine M.C., Dooley D.M., Freeman H.C., Guss J.M., Matsunami H.,
RA McIntire W.S., Ruggiero C.E., Tanizawa K., Yamaguchi H.;
RT "Crystal structures of the copper-containing amine oxidase from
RT Arthrobacter globiformis in the holo and apo forms: implications for
RT the biogenesis of topaquinine."
RL Biochemistry 36:16116-16133(1997).
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) -> RCHO + NH(3) +
CC H(2)O(2).
CC -1- COFACTOR: Binds 1 copper ion and 1 topaquinine per subunit.
CC -1- SUBUNIT: HOMODIMER.
CC -1- INDUCTION: BY PHENETHYLAMINE.
CC -1- PTM: Topaquinine (TPO) is generated by copper-dependent
CC autooxidation of a specific tyrosyl residue.
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAQUNINE OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U03517; AAA18114.1; -.
DR PDB: 1AVK; 18-MAR-98.
DR PDB: 1AVL; 18-MAR-98.
DR InterPro: IPR000269; CUNH_oxidase.
DR Pfam: PF01179; Cu_amine_oxid; 1.
DR Pfam: PF02727; Cu_amine_oxid2; 1.

```



```

0Y      87 KAAALAHND--RGSPPPREALAIYFF-----GROQOPNVSELVGLPBPSPYMRDVTYR 140
Db      38 RLAYGLVDLPARGAGSEADRRFRFHHIDVSGARQ-----EYTVSV 79
0Y      141 HGGPL-----PYHRRPVLFOEYLDIDDIHFN-----RELPOAS 173
Db      80 TNGVTIASAVELDTAATGCLPYLEEFEVEVEYOLLADDERWLKALAARNLDVSKYRAVAPLSA 139
0Y      174 GLHHCCFYKRRGRNLVMTTAPRGL-----QSGDRATWPGLYNNSGAGFFLHHNGLELL 229
Db      140 GVFEVA---EERGRKIL-----KGLAVYQDPEDESAM---AHPVDGLVAYDVVSKREY 187
0Y      230 VNHKALDPAKMTIOKVYOGARYYDSLAOLEAFEAGLVNVVLIIPNGTGGSWLSKSPVY 289
Db      188 ---RVIDGVFPVPA--EHNNTDP-----ELNGPLRT 215
0Y      290 GAPAPLOFYPOGPRPSVO--GSRVASLSMTYFSRGLDAFGSPRIFEDVRFO--GEUL---VEI 344
Db      216 TOKPSTIQPGPSTFYGVGHIEKWSLDVGFVREGVVLHNTAFRQDRLRPINRA 275
0Y      345 SLOEALAIYGGNSPAAMTTRYVD--GGFPMGXKTYTBLTRGVDC---PYLATYVDWHFLE 399
Db      276 SIAEKNVYVGPSPPIRSMQNYNFDTEGYLVGQVANSLELGGCGLGITYLSPIYSDAF--- 332
0Y      400 SQAPRTIDACVFEQONGLPLRRHSDLYSHYFSGLAETVLYVRSMTLLNYDVYMDTV 459
Db      333 -GNPEIENGICMEHEDDGLIAK--HSDLMGSIYTRRRNRVYIEFTTIGNVYDGFWY 389
0Y      460 FHPGSAIEFIREYATYI--SAFLFCATQKQNOVESEHTLGVHTSHAFHKVLDVDAULEN 518
Db      390 LYLDOTIEFEAKATVYVTSAPREGSGDNT--SQLPGLGAPRHOIIFSAKLDMAIDGFTN 448
0Y      519 VWMAEDMVFVPAVAVWSEBHOLORLQVTRKYLEMEQOAAFLVGSATPRXYLASNHS--NK 577
Db      449 RVEEDDVYRQTMPCNBERGNAFS---KRTVLTRESAVREMDATGRWIIISNESKRR 505
0Y      578 MCHPRGYRI---QMISFAGEPLPONSMAKRGFSWRYQATYQKKEEPPSSSV--NOND 633
Db      506 LNEPVGKYLHAHNOPTLLADP---GSSIAARRAAFTKDLMTYRYADDEREYFPGDGVNONS 562
0Y      634 PWAPTVDSDEINNETIAGKDLVAVANTAGFLHIAEADIPTNTVANGNGVFLFIRYPND 693
Db      563 GGAGGLPSV--IAORDDIDGODIVVHNTGTLHFRAVEDMP--IMVDYVIGFLKLRREGFFD 618
0Y      694 EDP 696
Db      619 RSP 621

```

DT 15-JUN-2002 [Rel. 41, Last annotation update]
DE amine oxidase [copper-containing] precursor (EC 1.4.3.6).
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophytes; Magnoliopsida; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
OX NCBI_TaxID:3888;
RN [1]
RP SEQUENCE FROM N.A..
RA Tipping A.J., McPherson M.J.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RC TISSUE=Seedling;
RX Medline:96396883; PubMed:8805580;
RA Kumar V., Dooley D.M., Freeman H.C., Guss J.M., Harvey I.,
RA McGill M.A., Wilce M.C., Zubak W.M.;
RT "Crystal structure of a eukaryotic (pea seedling) copper-containing
amine oxidase at 2.2-A resolution."
RL Structure 4:943-955(1996).
CC -1- CATALYTIC ACTIVITY: RCH(2)NH(2) + H(2)O + O(2) = RCHO + NH(3) +
H(2)O(2).
CC -1- COPROCTOL: Binds 1 copper ion, 1 manganese ion and 1 topaguinone
per subunit.
CC -1- SUBUNIT: HOMODIMER.
CC -1- PMW: Topaguinone (TPQ) is generated by copper-dependent
autooxidation of a specific tyrosyl residue (By similarity).
CC -1- SIMILARITY: BELONGS TO THE COPPER/TOPAGUINONE OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
DR EMBL, L39931; AAA62490.1; .
DR PDB: 1KSI; 24-DEC-97.
DR InterPro: IPR000269; CuNH_oxidase.
DR Pfam: PF011179; Cu_amine_oxid; 1.
DR Pfam: PF02727; Cu_amine_oxidn; 1.
DR Pfam: PF02728; Cu_amine_oxidn; 1.
DR PROSITE, PS01164; COPPER_AMINE_OXID_1; 1.
DR PROSITE, PS01165; COPPER_AMINE_OXID_2; FALSE_NEG.
DR Oxidoreductases; Metal-binding; Copper; Manganese; TPQ; glycoprotein;
KW Signal; 3D-structure.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 25 674 AMINE OXIDASE [COPPER-CONTAINING].
FT METAL 476 476 MANGANESE.
FT METAL 477 477 MANGANESE (VIA CARBONYL OXYGEN).
FT METAL 478 478 MANGANESE.
FT METAL 617 617 MANGANESE.
FT METAL 618 618 MANGANESE.
FT MOD_RES 412 412 MANGANESE (VIA CARBONYL OXYGEN).
FT METAL 467 467 TOPAGUINONE.
FT METAL 469 469 COPPER.
FT METAL 469 469 COPPER.
FT METAL 628 628 COPPER.
FT DISULFID 162 183
FT DISULFID 344 370
FT CARBOHYD 156 156 N-LINKED (GLCNAC. . .).
FT CARBOHYD 389 389 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 583 583 N-LINKED (GLCNAC. . .) (PROBABLE).
SQ SEQUENCE 674 AA; 76358 MW; 30735390071DD1BE CRC64;

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 07:58:08 ; Search time 18.4058 Seconds

(without alignments)
1578.232 Million cell updates/sec

Title: US-10-081-408-4

Perfect score: 1179
Sequence: 1 MSPILGKWKIKGLVOPTRLL.....IAMPLOQGNATFGGDHPK 218

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq.101002.*
1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1179	100.0	218	AAW84205	Protein comprising
2	1179	100.0	218	ABB09809	Glutathione-S-tran
3	1179	100.0	224	AAW02122	Glutathione-S-tran
4	1179	100.0	225	AAW02123	Glutathione-S-tran
5	1179	100.0	232	AAW72793	Glutathione-S-tran
6	1179	100.0	240	AAW73909	GST protein sequen
7	1179	100.0	245	AAW31612	Amino acid sequen
8	1179	100.0	247	AAW74114	GST/GI transport r
9	1179	100.0	247	AAW74120	GST/GI transport r
10	1179	100.0	247	AAW74103	GST/GI transport r

11	1179	100.0	248	AAW74112	GST/GI transport r
12	1179	100.0	248	AAW74113	GST/GI transport r
13	1179	100.0	248	AAW74121	GST/GI transport r
14	1179	100.0	248	AAW74122	GST/GI transport r
15	1179	100.0	248	AAW74100	GST/GI transport r
16	1179	100.0	248	AAW74101	GST/GI transport r
17	1179	100.0	248	AAW74102	GST/GI transport r
18	1179	100.0	249	AAW74115	GST/GI transport r
19	1179	100.0	257	AAW74111	GST/GI transport r
20	1179	100.0	257	AAW74119	GST/GI transport r
21	1179	100.0	257	AAW74124	GST/GI transport r
22	1179	100.0	257	AAW74126	GST/GI transport r
23	1179	100.0	257	AAW74105	GST/GI transport r
24	1179	100.0	257	AAW74109	GST/GI transport r
25	1179	100.0	258	AAW74117	GST/GI transport r
26	1179	100.0	258	AAW74104	GST/GI transport r
27	1179	100.0	259	AAW74130	GST/GI transport r
28	1179	100.0	259	AAW74118	GST/GI transport r
29	1179	100.0	259	AAW74125	GST/GI transport r
30	1179	100.0	259	AAW74110	GST/GI transport r
31	1179	100.0	262	AAW74128	GST/GI transport r
32	1179	100.0	264	AAW74129	GST/GI transport r
33	1179	100.0	267	AAW74106	GST/GI transport r
34	1179	100.0	271	AAW74103	GST/GI transport r
35	1179	100.0	272	AAW56327	Cap-A-B fusion pro
36	1179	100.0	272	AAW74116	Amino acid sequenc
37	1179	100.0	277	AAW74098	GST/GI transport r
38	1179	100.0	277	AAW74099	GST/GI transport r
39	1179	100.0	277	AAW74107	GST/GI transport r
40	1179	100.0	279	AAW74108	GST/GI transport r
41	1179	100.0	279	AAW74097	GST/GI transport r
42	1179	100.0	282	AAW56326	Amino acid sequenc
43	1179	100.0	282	AAW74127	GST/GI transport r
44	1179	100.0	282	AAW74123	GST/GI transport r
45	1179	100.0	282	AAW74095	GST/GI transport r

ALIGNMENTS

RESULT 1	
AAW84205	standard; Protein; 218 AA.
AC	AAW84205:
XX	
XX	25-MAR-1999 (first entry)
DE	Protein comprising a purification tag.
XX	
XX	Proteinase site; bone morphogenetic fusion protein; bone binding site;
KW	bone morphogenetic protein; transforming growth factor beta;
KW	active fragment; wound healing; bone growth; purification tag.
XX	
OS	Unidentified.
XX	
PN	WO9855137-A1.
XX	
PD	10-DEC-1998.
XX	
PF	02-JUN-1998: 98WO-US11189.
XX	
PR	03-JUN-1997: 97US-0868452.
XX	
PA	(HALL/) HALL F. L.
PA	(HAMB/) HAN B.
PA	(NIMK/) NIMKI M. E.
PA	(SHOR/) SHORS E. C.
PA	(WUL/) WU L.
PI	Hall FL, Han B, Nimki ME, Shors EC, Wu L;
XX	
DR	WPI: 1999-059875/05.

DR N-PSDB; AAV99373.
 XX
 PT New bone morphogenetic fusion proteins - comprising a purification
 tag and a bone morphogenetic active fragment, used for enhancing
 wound healing or bone growth
 PT
 XX
 PS Disclosure; Page 39-40; 64pp; English.
 CC
 CC The present sequence represents a protein comprising a purification tag
 CC that was used in the creation of the bone morphogenetic fusion proteins
 CC of the invention. The bone morphogenetic fusion protein may contain some
 CC or all of the following elements: a purification tag, a proteinase site,
 CC an ECM/bone binding site, a second proteinase site, and a bone
 CC morphogenetic protein active fragment. The fusion proteins of the
 CC invention also includes proteins that have transforming growth factor
 CC beta active fragments instead of bone morphogenetic protein active
 CC fragments. The bone morphogenetic fusion proteins can be used for
 CC enhancing wound healing or bone growth.
 CC
 SQ Sequence 218 AA:
 Query Match 100.0%; Score 1179; DB 20; Length 218;
 Best Local Similarity 100.0%; Pred. No. 4e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSPILGYWKIKGLVQPTRLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYID 60
 DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYID 60
 QY 61 GDVLTQSMATIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGSRIAYSKDFETLKV 120
 DB 61 GDVLTQSMATIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGSRIAYSKDFETLKV 120
 QY 121 DFLSKLPBEMKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVLYLMDPCLDAFPKLVCFK 180
 DB 121 DFLSKLPBEMKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVLYLMDPCLDAFPKLVCFK 180
 QY 181 KRIFAIPQIDKYLKSSKIYIAMPLOGMATFGGSDHPK 218
 DB 181 KRIFAIPQIDKYLKSSKIYIAMPLOGMATFGGSDHPK 218
 RESULT 2
 ABB09809
 ID ABB09809 standard; Protein; 218 AA.
 AC ABB09809;
 XX
 XX 22-JUL-2002 (first entry)
 DE Glutathione-S-transferase (GST) of Schistosoma japonicum.
 KW Glutathione-S-transferase; GST; enzyme; metallothionein; MT-IIA;
 heavy metal contact; UV light.
 KW
 XX Schistosoma japonicum.
 OS
 XX FR2813529-A1.
 PN
 XX 08-MAR-2002.
 PD
 XX 31-JUL-2001; 2001FR-0010239.
 PF
 XX 06-SEP-2000; 2000ES-0000217.
 PR
 XX (PROV-) PROVITAL SA.
 PA
 XX Armengol SR, Benalgués MA;
 PI
 XX WPI: 2002-332437/37.
 DR
 XX Topical compositions, used to protect against heavy metals and
 PT ultraviolet radiation, contain metallothionein proteins -

XX
 XX Example; Page 28-29; 33pp; French.
 PS
 CC The present sequence represents the glutathione-S-transferase (GST)
 CC protein of Schistosoma japonicum. A human metallothionein class IIA
 CC (MT-IIA) was expressed fused to the present GST protein, for
 CC purification purposes. The MT-IIA protein was used to produce
 CC topical pharmaceutical and cosmetic compositions, in the form of gels,
 CC creams, ointments, soaps and body lotions. In the compositions, the
 CC metallothionein proteins have a prosthetic group attached to a zinc
 CC ion. The topical compositions can be used for protection of human or
 CC animal external tissue against the effects of heavy metal contact and
 UV light.
 CC
 SQ Sequence 218 AA:
 Query Match 100.0%; Score 1179; DB 23; Length 218;
 Best Local Similarity 100.0%; Pred. No. 4e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSPILGYWKIKGLVQPTRLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYID 60
 DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKEEHLERDEGDKRNKKFELGLEFPNLPYYID 60
 QY 61 GDVLTQSMATIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGSRIAYSKDFETLKV 120
 DB 61 GDVLTQSMATIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGSRIAYSKDFETLKV 120
 QY 121 DFLSKLPBEMKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVLYLMDPCLDAFPKLVCFK 180
 DB 121 DFLSKLPBEMKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVLYLMDPCLDAFPKLVCFK 180
 QY 181 KRIFAIPQIDKYLKSSKIYIAMPLOGMATFGGSDHPK 218
 DB 181 KRIFAIPQIDKYLKSSKIYIAMPLOGMATFGGSDHPK 218
 RESULT 3
 AAM02122
 ID AAM02122 standard; Protein; 224 AA.
 XX AAM02122;
 AC
 XX 28-OCT-1996 (first entry)
 DT
 XX Glutathione-s-transferase epitope.
 DE
 XX Bone resorption disease; osteoporosis; src SH2 domain antagonist;
 KW src homology 2 domain; glutathione-s-transferase; Tyr SH2;
 KW Grb2 SH2; SH-PTP2 SH2.
 KW
 XX Not specified.
 OS
 XX EP727211-A1.
 PN
 XX 21-AUG-1996.
 PD
 XX 07-FEB-1996; 96EP-0200270.
 PE
 XX 29-DEC-1995; 95US-0580868.
 PR 10-FEB-1995; 95US-0386381.
 PR 07-MAR-1995; 95US-0400220.
 PR 30-JUN-1995; 95US-0497357.
 PR 11-OCT-1995; 95US-0541080.
 XX
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 PA
 XX Dunnington DJ;
 PI
 XX WPI: 1996-372674/38.
 DR
 XX Use of selective src SH2 domain ligand - to prepare medicament for
 PT treating bone resorption disease

XX Example 11: Page 31-32; 47pp; English.

PS A glutathione-s-transferase epitope (AAW02122) was used to construct
 CC fusion proteins with human SH2 domains from fyn (see also AAW02124),
 CC Grb2 (AAW02127) and SH-PP2 (AAW02125); a similar epitope (AAW02123) was
 CC used to construct a fusion protein with human p85 SH2 domain (see
 CC also AAW02126). These fusion proteins, and tagged proteins including
 CC other SH2 domains (see also AAW02119-21), can be used in binding
 CC assays to determine the specificity of cpds. to inhibit SH2 domains;
 CC cpds. that selectively inhibit the human src SH2 domain are useful
 CC in treating bone resorption diseases such as osteoporosis.

XX Sequence 224 AA:

SO Query Match 100.0%; Score 1179; DB 17; Length 224;
 Best Local Similarity 100.0%; Pred. No. 4,1e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSPTLGWTKIGLVOPRLLEYLEEKEEHLERDGDGKRNKKFELGLEFPNLPYYID 60
 DB 1 MSPILGWTKIGLVOPRLLEYLEEKEEHLERDGDGKRNKKFELGLEFPNLPYYID 60
 OY 61 GDVKLTQSMATIRYIADKHNMLGCGPKERAIEISMLEGAVLDIRYGSRIAYSKDFETLV 120
 DB 61 GDVKLTQSMATIRYIADKHNMLGCGPKERAIEISMLEGAVLDIRYGSRIAYSKDFETLV 120
 OY 121 DFLSKLPKMEKMFEDRLCHKTYLNGDVTHTPDMFLYDALDVLYMDPCLDAFPKLVCFK 180
 DB 121 DFLSKLPKMEKMFEDRLCHKTYLNGDVTHTPDMFLYDALDVLYMDPCLDAFPKLVCFK 180
 OY 121 DFLSKLPKMEKMFEDRLCHKTYLNGDVTHTPDMFLYDALDVLYMDPCLDAFPKLVCFK 180
 DB 121 DFLSKLPKMEKMFEDRLCHKTYLNGDVTHTPDMFLYDALDVLYMDPCLDAFPKLVCFK 180
 OY 181 KRTEAIPOIDKYLKSSKYIAMPLOGQATFGGSDHPK 218
 DB 181 KRTEAIPOIDKYLKSSKYIAMPLOGQATFGGSDHPK 218

Db 181 KRTEAIPOIDKYLKSSKYIAMPLOGQATFGGSDHPK 218

RESULT 4
 AAW02123
 ID AAW02123 standard; Protein; 225 AA.

XX AAW02123:
 AC AAW02123:
 XX 28-OCT-1996 (first entry)
 DT 28-OCT-1996 (first entry)
 DE Glutathione-s-transferase epitope.
 XX Bone resorption disease; osteoporosis; src SH2 domain antagonist;
 KW src homology 2 domain; glutathione-s-transferase; p85 SH2.
 XX Not specified.
 OS Not specified.
 XX EP727211-A1.
 PN EP727211-A1.
 XX 21-AUG-1996.
 PD 21-AUG-1996.
 XX 07-FEB-1996; 96EP-0200270.
 PF 07-FEB-1996; 96EP-0200270.
 XX 29-DEC-1995; 95US-0580868.
 PR 10-FEB-1995; 95US-0386381.
 PR 07-MAR-1995; 95US-0400220.
 PR 30-JUN-1995; 95US-0497357.
 PR 11-OCT-1995; 95US-0541080.
 XX (SMIR) SMITHKLINE BEECHAM CORP.
 PA (SMIR) SMITHKLINE BEECHAM CORP.
 XX Dunnington DJ;
 PI Dunnington DJ;
 XX WPI; 1996-372674/38.
 DR WPI; 1996-372674/38.
 XX Use of selective src SH2 domain ligand - to prepare medicament for
 PT treating bone resorption disease
 XX Example 11: Page 31-32; 47pp; English.

XX Example 11: Page 31-32; 47pp; English.

PS A glutathione-s-transferase epitope (AAW02122) was used to construct
 CC fusion proteins with human SH2 domains from fyn (see also AAW02124),
 CC Grb2 (AAW02127) and SH-PP2 (AAW02125); a similar epitope (AAW02123) was
 CC used to construct a fusion protein with human p85 SH2 domain (see
 CC also AAW02126). These fusion proteins, and tagged proteins including
 CC other SH2 domains (see also AAW02119-21), can be used in binding
 CC assays to determine the specificity of cpds. to inhibit SH2 domains;
 CC cpds. that selectively inhibit the human src SH2 domain are useful
 CC in treating bone resorption diseases such as osteoporosis.

XX Sequence 225 AA:

SO Query Match 100.0%; Score 1179; DB 17; Length 225;
 Best Local Similarity 100.0%; Pred. No. 4,2e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSPTLGWTKIGLVOPRLLEYLEEKEEHLERDGDGKRNKKFELGLEFPNLPYYID 60
 DB 1 MSPILGWTKIGLVOPRLLEYLEEKEEHLERDGDGKRNKKFELGLEFPNLPYYID 60
 OY 61 GDVKLTQSMATIRYIADKHNMLGCGPKERAIEISMLEGAVLDIRYGSRIAYSKDFETLV 120
 DB 61 GDVKLTQSMATIRYIADKHNMLGCGPKERAIEISMLEGAVLDIRYGSRIAYSKDFETLV 120
 OY 121 DFLSKLPKMEKMFEDRLCHKTYLNGDVTHTPDMFLYDALDVLYMDPCLDAFPKLVCFK 180
 DB 121 DFLSKLPKMEKMFEDRLCHKTYLNGDVTHTPDMFLYDALDVLYMDPCLDAFPKLVCFK 180
 OY 121 DFLSKLPKMEKMFEDRLCHKTYLNGDVTHTPDMFLYDALDVLYMDPCLDAFPKLVCFK 180
 DB 121 DFLSKLPKMEKMFEDRLCHKTYLNGDVTHTPDMFLYDALDVLYMDPCLDAFPKLVCFK 180
 OY 181 KRTEAIPOIDKYLKSSKYIAMPLOGQATFGGSDHPK 218
 DB 181 KRTEAIPOIDKYLKSSKYIAMPLOGQATFGGSDHPK 218

Db 181 KRTEAIPOIDKYLKSSKYIAMPLOGQATFGGSDHPK 218

RESULT 5
 AAR72793
 ID AAR72793 standard; Protein; 232 AA.

XX AAR72793:
 AC AAR72793:
 XX 16-OCT-1995 (first entry)
 DT 16-OCT-1995 (first entry)
 DE Glutathione-s-transferase.
 XX SOD-1; senescent cell-derived inhibitor; DNA synthesis;
 KW senescence; cell proliferation; cancer; therapeutic; vulnary;
 KW fusion protein; glutathione-s-transferase.
 XX Schistosoma japonicum.
 OS Schistosoma japonicum.
 XX WO9506415-A.
 PN WO9506415-A.
 XX 09-MAR-1995.
 PD 09-MAR-1995.
 XX 26-AUG-1994; 94WO-US09700.
 PF 26-AUG-1994; 94WO-US09700.
 XX 13-JUL-1994; 94US-0274535.
 PR 30-AUG-1993; 93US-0113372.
 PR 17-NOV-1993; 93US-0153564.
 PR 03-JAN-1994; 94US-0160814.
 PR 25-FEB-1994; 94US-0203535.
 PR 15-APR-1994; 94US-0229420.
 PR 30-JUN-1994; 94US-0268439.
 XX (BAYU) BAYLOR COLLEGE MEDICINE.
 PA (BAYU) BAYLOR COLLEGE MEDICINE.
 XX Smith JR;
 PI Smith JR;
 XX WPI; 1995-131101/17.
 DR WPI; 1995-131101/17.
 XX N-PSDB; AAG86774.
 DR N-PSDB; AAG86774.
 XX Nucleic acid encoding a protein or polypeptide that inhibits DNA
 PT synthesis in a recipient cell - useful to inhibit cell

PT proliferation in tumour cells, treat wound or burn tissue, or as
PT an antiviral or antiparasitic agent
XX
XX
PS Disclosure; Page 130; 169pp; English.
XX
CC The senescent cell-derived inhibitor-1 (SDI-1) inhibits DNA
CC synthesis in a recipient cell and is capable of associating with a
CC cyclin (esp. cyclin D1) or cyclin-dependent kinase (esp. CDK2).
CC Production of a fusion protein of SDI-1 and glutathione-S-transferase
CC (AAR72793) facilitated the preparation and isolation of recombinant
CC SDI-1.
CC
XX
SQ Sequence 232 AA;
Query Match 100.0%; Score 1179; DB 16; Length 232;
Best Local Similarity 100.0%; Pred. No. 4.3e-113;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEERDEGDKRNKKFELGLEFPNLPYYID 60
QY 61 GDVKTQSMALIRYIADKHNMLGCGPKERAEISMLEGAVLDIRGVSRIVASKDFETLKV 120
DB 61 GDVKTQSMALIRYIADKHNMLGCGPKERAEISMLEGAVLDIRGVSRIVASKDFETLKV 120
QY 121 DFLSKLPKMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPNCCLDAFPKLVCFK 180
DB 121 DFLSKLPKMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPNCCLDAFPKLVCFK 180
QY 181 KRIEAIPOIDKYLKSSKYIAMPLQGWQATFGGDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKYIAMPLQGWQATFGGDHPPK 218
RESULT 6
ID AAW73909 standard; Protein; 240 AA.
AC AAW73909;
XX
XX AAW73909;
XX
DT 13-APR-1999 (first entry)
DE GST protein sequence.
XX
XX Cell surface nonexpressive functional protein; cell-permeating protein;
KM protein screening; protein production; GST; glutathione-S-transferase.
XX
XX Schistosoma japonicum.
OS
XX JP11009278-A.
PN
XX 19-JAN-1999.
PD
XX 23-JUN-1997; 97JP-0165788.
PF
XX 23-JUN-1997; 97JP-0165788.
PR
XX 23-JUN-1997; 97JP-0165788.
PA
XX (SUME) SUMITOMO ELECTRIC IND CO.
XX
XX WPI; 1999-145895/13.
DR
XX N-PSDB; AAX01312.
PT
XX Microbe having a nonexpressive functional protein on its surface -
PT for screening and large-scale preparation of a selected function
PT protein
XX
XX Disclosure; Page 8-9; 13pp; Japanese.
PS
XX This sequence is the glutathione-S-transferase (GST) protein, which
CC was expressed in the microbe of the invention. The microbe has a cell
CC surface nonexpressive functional protein on its surface, fused to a
CC support protein derived from a cell-permeating protein. The microbe can

CC be used for screening for biological activity of a cell surface
CC nonexpressive protein. It can also be used for preparation of a selected
CC functional protein in a large amounts.
XX
XX
SQ Sequence 240 AA;
Query Match 100.0%; Score 1179; DB 20; Length 240;
Best Local Similarity 100.0%; Pred. No. 4.6e-113;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEERDEGDKRNKKFELGLEFPNLPYYID 60
QY 61 GDVKTQSMALIRYIADKHNMLGCGPKERAEISMLEGAVLDIRGVSRIVASKDFETLKV 120
DB 61 GDVKTQSMALIRYIADKHNMLGCGPKERAEISMLEGAVLDIRGVSRIVASKDFETLKV 120
QY 121 DFLSKLPKMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPNCCLDAFPKLVCFK 180
DB 121 DFLSKLPKMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPNCCLDAFPKLVCFK 180
QY 181 KRIEAIPOIDKYLKSSKYIAMPLQGWQATFGGDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKYIAMPLQGWQATFGGDHPPK 218
RESULT 7
ID AAB31612 standard; Protein; 245 AA.
AC AAB31612;
XX
XX AAB31612;
XX
DT 30-APR-2001 (first entry)
DE Amino acid sequence of glutathione-S-transferase (GST).
XX
XX Heat shock protein; Hsp; Th1 response; Th1 cell; CD4+ T lymphocyte cell;
KW lymphocyte; Hsp65; Hsp40; Hsp10; Hsp60; Hsp71; microbial pathogen;
KW glutathione-S-transferase; GST.
XX
XX Unidentified.
OS
XX WO200104344-A2.
PN
XX 18-JAN-2001.
PD
XX 10-JUL-2000; 2000WO-US18828.
PF
XX 08-JUL-1999; 99US-0143757.
PR
XX (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
PA
XX Siegel M, Chu NR, Mizen LA;
PI
XX WPI; 2001-138361/14.
DR
XX N-PSDB; AAF25015.
PT
XX Screening for compounds that stimulate Th1-like responses in CD4+ T
PT lymphocyte cells -
XX
XX Example 9; Fig 8; 88pp; English.
XX
CC The present sequence represents a glutathione-S-transferase (GST), which
CC was used linked to HPV16 E7 protein, for purification purposes. The E7
CC protein was used to construct a fusion protein with Mycobacterium bovis
CC heat shock proteins (Hsp). The fusion proteins are used in the method of
CC the invention. The specification describes a method of determining
CC whether a compound stimulates a Th1-like response. Th1 cells are a subset
CC of CD4+ T lymphocyte cells. The method comprises contacting native
CC lymphocytes in vitro with a fusion protein comprising at least a fragment
CC of Hsp, and then detecting the Th1-like response exhibited by the cell
CC sample. The proteins which may be used in the method of the invention are

CC Hsp65, Hsp40, Hsp10, Hsp60, and Hsp71. The method may be used to identify
 CC compounds that stimulate Th1-like responses in response to microbial
 CC pathogens.
 CC
 XX
 SQ Sequence 245 AA;
 Query Match 100.0%; Score 1179; DB 22; Length 245;
 Best Local Similarity 100.0%; Pred. No. 4.7e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSPILGYWKIKGLVOPTRLLLEYLEEYEEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60
 DB 1 MSPILGYWKIKGLVOPTRLLLEYLEEYEEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60
 QY 61 GDVKLTQSMATIRYIADKHNMLGCGPKERAISMLGAVLDIRGVSRIVSKDFETLKV 120
 DB 61 GDVKLTQSMATIRYIADKHNMLGCGPKERAISMLGAVLDIRGVSRIVSKDFETLKV 120
 QY 121 DFLSKLPENLKMFEEDRLCHHTYTLNGDHVTHPDMFLYDALDVLYMDPMCLDAPFKLYCFK 180
 DB 121 DFLSKLPENLKMFEEDRLCHHTYTLNGDHVTHPDMFLYDALDVLYMDPMCLDAPFKLYCFK 180
 QY 181 KRIEAIPOIDKYLKSSKYIAMPLOGWQATGEGGDHPK 218
 DB 181 KRIEAIPOIDKYLKSSKYIAMPLOGWQATGEGGDHPK 218
 RESULT 8
 AAW74114
 ID AAW74114 standard; peptide: 247 AA.
 AC AAW74114;
 XX
 DT 04-MAY-1999 (first entry)
 XX
 DE GST/GI transport receptor binding protein fusion sequence.
 XX
 KM Gastro-intestinal transport receptor; binding protein; hsi; HPT1;
 KM D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KM intestinal peptide-associated transporter; hypertension; diabetes;
 KM osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KM therapeutic agent delivery; therapy; fusion protein.
 XX
 OS Homo sapiens.
 XX
 PN WO9851325-A2.
 XX
 PD 19-NOV-1998.
 XX
 PF 15-MAY-1998; 98WO-US10088.
 XX
 PR 15-MAY-1997; 97US-0046595.
 XX
 PA (CYTO-) CYTOGEN CORP.
 PA (ELAN-) ELAN CORP PLC.
 XX
 PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ;
 PI Mahony DJ, Patterson CA, Singleton J;
 DR WPI: 1999-009568/01.
 XX
 PT New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide
 XX
 PS Example: Page 217; 294pp; English.
 CC This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (i) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (HPT1),

CC hPEPT1, D2H and human sucrose-isomaltase complex (hsi). (i) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid. Imaging agents
 CC and antigens. (ii) may also provide targeting to the GI tract. Other uses
 CC of (i) are: (1) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (i).
 CC Immunogenic analogues or derivatives of (i) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (i), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.
 XX
 SQ Sequence 247 AA;
 Query Match 100.0%; Score 1179; DB 20; Length 247;
 Best Local Similarity 100.0%; Pred. No. 4.7e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSPILGYWKIKGLVOPTRLLLEYLEEYEEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60
 DB 1 MSPILGYWKIKGLVOPTRLLLEYLEEYEEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60
 QY 61 GDVKLTQSMATIRYIADKHNMLGCGPKERAISMLGAVLDIRGVSRIVSKDFETLKV 120
 DB 61 GDVKLTQSMATIRYIADKHNMLGCGPKERAISMLGAVLDIRGVSRIVSKDFETLKV 120
 QY 121 DFLSKLPENLKMFEEDRLCHHTYTLNGDHVTHPDMFLYDALDVLYMDPMCLDAPFKLYCFK 180
 DB 121 DFLSKLPENLKMFEEDRLCHHTYTLNGDHVTHPDMFLYDALDVLYMDPMCLDAPFKLYCFK 180
 QY 181 KRIEAIPOIDKYLKSSKYIAMPLOGWQATGEGGDHPK 218
 DB 181 KRIEAIPOIDKYLKSSKYIAMPLOGWQATGEGGDHPK 218
 RESULT 9
 AAW74120
 ID AAW74120 standard; peptide: 247 AA.
 AC AAW74120;
 XX
 DT 04-MAY-1999 (first entry)
 XX
 DE GST/GI transport receptor binding protein fusion sequence.
 XX
 KM Gastro-intestinal transport receptor; binding protein; hsi; HPT1;
 KM D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KM intestinal peptide-associated transporter; hypertension; diabetes;
 KM osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KM therapeutic agent delivery; therapy; fusion protein.
 XX
 OS Homo sapiens.
 XX
 PN WO9851325-A2.
 XX
 PD 19-NOV-1998.
 XX
 PF 15-MAY-1998; 98WO-US10088.
 XX
 PR 15-MAY-1997; 97US-0046595.
 XX
 PA (CYTO-) CYTOGEN CORP.
 PA (ELAN-) ELAN CORP PLC.
 XX
 PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ;
 PI Mahony DJ, Patterson CA, Singleton J;
 DR WPI: 1999-009568/01.
 XX
 PT New proteins that bind specifically to receptors in the

PT gastro-intestinal tract and related nucleic acid - chimaeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastro-intestinal tract, e.g. insulin or leuprolide
 XX
 XX
 PS Example: Page 221; 294pp; English.

CC This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (HPT1),
 CC hPPT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.
 CC
 XX

SO Sequence 247 AA;

Query Match 100.0%; Score 1179; DB 20; Length 247;
 Best Local Similarity 100.0%; Pred. No. 4.7e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSPILGKWKIKGLVPTRLLEYLEEKYEENHLYERDEGDKWKKFELGLEPNNLPYYID 60
 DB 1 MSPILGKWKIKGLVPTRLLEYLEEKYEENHLYERDEGDKWKKFELGLEPNNLPYYID 60
 OY 61 GGVKLTQSMALIRYIADKHNMLGCGPKERAETSMLEGAVALDIRYVSRATSKDEFTLV 120
 DB 61 GGVKLTQSMALIRYIADKHNMLGCGPKERAETSMLEGAVALDIRYVSRATSKDEFTLV 120
 OY 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHYTHDPFMLYDALDVLYNDPMLDAFPKLVCFK 180
 DB 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHYTHDPFMLYDALDVLYNDPMLDAFPKLVCFK 180
 OY 181 KRIEAIPOIDKYLKSSKYIAMPLOGMOATFGGDDHPK 218
 DB 181 KRIEAIPOIDKYLKSSKYIAMPLOGMOATFGGDDHPK 218

RESULT 10

ID AAM74103 standard; peptide: 247 AA.

AC AAM74103;

DT 04-MAY-1999 (first entry)

DE GST/GI transport receptor binding protein fusion sequence.

KW Gastro-intestinal transport receptor; binding protein; hSI; HPT1;
 KW D2H; hPPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KW intestinal peptide-associated transporter; hypertension; diabetes;
 KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KW therapeutic agent delivery; therapy; fusion protein.

OS Homo sapiens.

PN W09851325-A2.

XX 19-NOV-1998.

PF 15-MAY-1998; 98WO-US10088.

XX

PR 15-MAY-1997; 97US-0046595.
 XX
 XX (CYTO-) CYTOGEN CORP.
 PA (ELAN-) ELAN CORP PLC.
 XX
 XX
 PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin JI;
 PI Omanony DJ, Patterson CA, Singleton J;
 DR WPI, 1999-009568/01.

PT New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimaeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastro-intestinal tract, e.g. insulin or leuprolide
 XX
 XX
 PS Example: Page 209; 294pp; English.

CC This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (HPT1),
 CC hPPT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.
 CC
 XX

SO Sequence 247 AA;

Query Match 100.0%; Score 1179; DB 20; Length 247;
 Best Local Similarity 100.0%; Pred. No. 4.7e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSPILGKWKIKGLVPTRLLEYLEEKYEENHLYERDEGDKWKKFELGLEPNNLPYYID 60
 DB 1 MSPILGKWKIKGLVPTRLLEYLEEKYEENHLYERDEGDKWKKFELGLEPNNLPYYID 60
 OY 61 GGVKLTQSMALIRYIADKHNMLGCGPKERAETSMLEGAVALDIRYVSRATSKDEFTLV 120
 DB 61 GGVKLTQSMALIRYIADKHNMLGCGPKERAETSMLEGAVALDIRYVSRATSKDEFTLV 120
 OY 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHYTHDPFMLYDALDVLYNDPMLDAFPKLVCFK 180
 DB 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHYTHDPFMLYDALDVLYNDPMLDAFPKLVCFK 180
 OY 181 KRIEAIPOIDKYLKSSKYIAMPLOGMOATFGGDDHPK 218
 DB 181 KRIEAIPOIDKYLKSSKYIAMPLOGMOATFGGDDHPK 218

RESULT 11

ID AAM74112 standard; peptide: 248 AA.

AC AAM74112;

DT 04-MAY-1999 (first entry)

DE GST/GI transport receptor binding protein fusion sequence.

KW Gastro-intestinal transport receptor; binding protein; hSI; HPT1;
 KW D2H; hPPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KW intestinal peptide-associated transporter; hypertension; diabetes;

KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KM therapeutic agent delivery; therapy; fusion protein.
 XX Homo sapiens.
 OS MO9851325-A2.
 PN 19-NOV-1998.
 PD 15-MAY-1998; 98WO-US10088.
 XX PF 15-MAY-1997; 97US-0046595.
 XX PR 15-MAY-1997; 97US-0046595.
 XX PA (CYTO-) CYTOGEN CORP.
 XX PA (ELAN-) ELAN CORP PLC.
 XX PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ;
 XX PI Omahony DJ, Patterson CA, Singleton J;
 DR WPI: 1999-009568/01.
 XX New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimaeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide
 XX Example: Page 215-216; 294pp; English.
 PS This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (hPrl),
 CC hPrl, D2H and human sucrose-isomaltase complex (hSI). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.
 CC XX
 SO Sequence 248 AA:
 Query Match 100.0%; Score 1179; DB 20; Length 248;
 Best Local Similarity 100.0%; Pred. No. 4.8e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEHLERDEGDKWRKKFELGLEFPNLPYYID 60
 DB 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEHLERDEGDKWRKKFELGLEFPNLPYYID 60
 OY 61 GGVKLTQSMALIRYIADKHNMLGCGPKERAISMLGAVLDIRYGVSRIRAYSKDFETLKY 120
 DB 61 GGVKLTQSMALIRYIADKHNMLGCGPKERAISMLGAVLDIRYGVSRIRAYSKDFETLKY 120
 OY 121 DFLSKLPKMKFEDRLCKRTYLVNGDHVTHPDMFLYDALDVLYMDPCIDAFPKLVCFK 180
 DB 121 DFLSKLPKMKFEDRLCKRTYLVNGDHVTHPDMFLYDALDVLYMDPCIDAFPKLVCFK 180
 OY 181 KRTEALPOIDKYLKSSKSIYAMPLOGMCAATGGGDDHPK 218
 DB 181 KRTEALPOIDKYLKSSKSIYAMPLOGMCAATGGGDDHPK 218

ID AAM74113 standard; peptide: 248 AA.
 XX AAM74113:
 AC AAM74113:
 XX 04-MAY-1999 (first entry)
 DT GST/GI transport receptor binding protein fusion sequence.
 XX GST/GI transport receptor binding protein fusion sequence.
 DE Gastro-intestinal transport receptor; binding protein; hSI; hPrl;
 KM D2H; hPrl; human; GI tract receptor; sucrose-isomaltase complex;
 KM intestinal peptide-associated transporter; hypertension; diabetes;
 KM osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KM therapeutic agent delivery; therapy; fusion protein.
 XX Homo sapiens.
 OS MO9851325-A2.
 PN 19-NOV-1998.
 PD 15-MAY-1998; 98WO-US10088.
 XX PF 15-MAY-1997; 97US-0046595.
 XX PR 15-MAY-1997; 97US-0046595.
 XX PA (CYTO-) CYTOGEN CORP.
 XX PA (ELAN-) ELAN CORP PLC.
 XX PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ;
 XX PI Omahony DJ, Patterson CA, Singleton J;
 DR WPI: 1999-009568/01.
 XX New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimaeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide
 XX Example: Page 216; 294pp; English.
 PS This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (hPrl),
 CC hPrl, D2H and human sucrose-isomaltase complex (hSI). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.
 CC XX
 SO Sequence 248 AA:
 Query Match 100.0%; Score 1179; DB 20; Length 248;
 Best Local Similarity 100.0%; Pred. No. 4.8e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEHLERDEGDKWRKKFELGLEFPNLPYYID 60
 DB 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEHLERDEGDKWRKKFELGLEFPNLPYYID 60
 OY 61 GGVKLTQSMALIRYIADKHNMLGCGPKERAISMLGAVLDIRYGVSRIRAYSKDFETLKY 120
 DB 61 GGVKLTQSMALIRYIADKHNMLGCGPKERAISMLGAVLDIRYGVSRIRAYSKDFETLKY 120

QY 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLVYMDPMLDAFPKLVCFK 180
 DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLVYMDPMLDAFPKLVCFK 180
 QY 181 KRIEAIPOIDKYLKSSKXIAMPLOGMQATFGGDDHPK 218
 DB 181 KRIEAIPOIDKYLKSSKXIAMPLOGMQATFGGDDHPK 218

RESULT 13

AAW74121
 ID AAW74121 standard; peptide: 248 AA.

AAW74121;
 AC AAW74121;

DT 04-MAY-1999 (first entry)
 DE GST/GI transport receptor binding protein fusion sequence.

XX Gastro-intestinal transport receptor; binding protein; hSI; HPT1;
 KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KW intestinal peptide-associated transporter; hypertension; diabetes;
 KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KW therapeutic agent delivery; therapy; fusion protein.

XX Homo sapiens.

OS W09851325-A2.

PN 19-NOV-1998.

PD 15-MAY-1998; 98WO-US10088.

PF 15-MAY-1997; 97US-0046595.

PR (CYTO-) CYTOGEN CORP.

PA (ELAN-) ELAN CORP PLC.

PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ;
 PI Omahony DJ, Patterson CA, Singleton J;

DR WPI; 1999-009568/01.

PT New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide
 XX Example; Page 222; 294pp; English.

XX This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (HPT1),
 CC hPEPT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.

XX Sequence 248 AA;

Query Match 100.0%; Score 1179; DB 20; Length 248;

Best Local Similarity 100.0%; Pred. No. 4 8e-113;
 Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGKIKGLVOPTRILLEYLEEKYEHLPERDEGDKRNKKFELGLEFPNLPYYID 60
 DB 1 MSPILGYMKIKGLVOPTRILLEYLEEKYEHLPERDEGDKRNKKFELGLEFPNLPYYID 60
 QY 61 GDVKLTQSMATIRYADHNHMLGGCPKERAETSMLEGAVALDIRGVSRINASKDEETLKV 120
 DB 61 GDVKLTQSMATIRYADHNHMLGGCPKERAETSMLEGAVALDIRGVSRINASKDEETLKV 120
 QY 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLVYMDPMLDAFPKLVCFK 180
 DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLVYMDPMLDAFPKLVCFK 180
 QY 181 KRIEAIPOIDKYLKSSKXIAMPLOGMQATFGGDDHPK 218
 DB 181 KRIEAIPOIDKYLKSSKXIAMPLOGMQATFGGDDHPK 218

RESULT 14

AAW74122
 ID AAW74122 standard; peptide: 248 AA.

AAW74122;
 AC AAW74122;

DT 04-MAY-1999 (first entry)
 DE GST/GI transport receptor binding protein fusion sequence.

XX Gastro-intestinal transport receptor; binding protein; hSI; HPT1;
 KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KW intestinal peptide-associated transporter; hypertension; diabetes;
 KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KW therapeutic agent delivery; therapy; fusion protein.

XX Homo sapiens.

OS W09851325-A2.

PN 19-NOV-1998.

PD 15-MAY-1998; 98WO-US10088.

PF 15-MAY-1997; 97US-0046595.

PR (CYTO-) CYTOGEN CORP.

PA (ELAN-) ELAN CORP PLC.

PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ;
 PI Omahony DJ, Patterson CA, Singleton J;

DR WPI; 1999-009568/01.

PT New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide
 XX Example; Page 222-223; 294pp; English.

XX This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (HPT1),
 CC hPEPT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents

CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (1) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (11) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.

XX Sequence 248 AA;

Query Match 100.0%; Score 1179; DB 20; Length 248;

Best Local Similarity 100.0%; Pred. No. 4.8e-113; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVOPRLLEYLEEYEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60
 DB 1 MSPILGYWKIKGLVOPRLLEYLEEYEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60
 QY 61 GDVKLTOSMAIIRYADKHNMLGCGPKERAISMLEGAVLDIRGVSRIVASDFETLKY 120
 DB 61 GDVKLTOSMAIIRYADKHNMLGCGPKERAISMLEGAVLDIRGVSRIVASDFETLKY 120
 QY 121 DFLSKLPKEMLKMFEDRLCHRTYTLNGDVTHTPDEFLYDALDVLYMDPCLDAEPKLVCFK 180
 DB 121 DFLSKLPKEMLKMFEDRLCHRTYTLNGDVTHTPDEFLYDALDVLYMDPCLDAEPKLVCFK 180
 QY 181 KRIEAIPOIDKYLKSSKRYIAMPLOGWQATFGGDDHPRK 218
 DB 181 KRIEAIPOIDKYLKSSKRYIAMPLOGWQATFGGDDHPRK 218

RESULT 15

AAW74100

AAW74100

AAW74100

XX 04-MAY-1999 (first entry)
 DE GST/GI transport receptor binding protein fusion sequence.
 XX
 KM Gastro-intestinal transport receptor; binding protein; hSI; hPT1;
 KM DH: hPEPT1; human: GI tract receptor; sucrose-isomaltase complex;
 KM intestinal peptide-associated transporter; hypertension; diabetes;
 KM osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KM therapeutic agent delivery; therapy; fusion protein.

OS Homo sapiens.

PN W09851325-A2.

PD 19-NOV-1998.

PF 15-MAY-1998; 98MO-US10088.

PR 15-MAY-1997; 97US-0046595.

PA (CYTO-) CYTOGEN CORP.
 PA (ELAN-) ELAN CORP PLC.

PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin JF;
 PI Mahony DJ, Patterson CA, Singleton J;

DR WPI; 1999-009568/01.

XX New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide
 XX
 PS Example: Page 207; 294pp; English.
 XX This sequence represents a fusion protein between

CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (hPT1),
 CC hPEPT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (1) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (11) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.

XX Sequence 248 AA;

Query Match 100.0%; Score 1179; DB 20; Length 248;

Best Local Similarity 100.0%; Pred. No. 4.8e-113; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVOPRLLEYLEEYEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60
 DB 1 MSPILGYWKIKGLVOPRLLEYLEEYEHLYERDEGDKWRNKKFELGLEFPNLPYYID 60
 QY 61 GDVKLTOSMAIIRYADKHNMLGCGPKERAISMLEGAVLDIRGVSRIVASDFETLKY 120
 DB 61 GDVKLTOSMAIIRYADKHNMLGCGPKERAISMLEGAVLDIRGVSRIVASDFETLKY 120
 QY 121 DFLSKLPKEMLKMFEDRLCHRTYTLNGDVTHTPDEFLYDALDVLYMDPCLDAEPKLVCFK 180
 DB 121 DFLSKLPKEMLKMFEDRLCHRTYTLNGDVTHTPDEFLYDALDVLYMDPCLDAEPKLVCFK 180
 QY 181 KRIEAIPOIDKYLKSSKRYIAMPLOGWQATFGGDDHPRK 218
 DB 181 KRIEAIPOIDKYLKSSKRYIAMPLOGWQATFGGDDHPRK 218

Search completed: May 20, 2003, 11:33:54
 Job time : 18.4058 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:30:54 ; Search time 7.02587 Seconds

(without alignments)
912.939 Million cell updates/sec

Title: US-10-081-408-4

Perfect score: 1179
Sequence: 1 MSPILGYWKIKGLVQPTRL.....IANPLOGMQATFGGDDHPPK 218Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/2/1aa/5C.COMB.pep.*
- 4: /cgn2_6/ptodata/2/1aa/5D.COMB.pep.*
- 5: /cgn2_6/ptodata/2/1aa/5E.COMB.pep.*
- 6: /cgn2_6/ptodata/2/1aa/5F.COMB.pep.*
- 7: /cgn2_6/ptodata/2/1aa/5G.COMB.pep.*
- 8: /cgn2_6/ptodata/2/1aa/5H.COMB.pep.*
- 9: /cgn2_6/ptodata/2/1aa/5I.COMB.pep.*
- 10: /cgn2_6/ptodata/2/1aa/5J.COMB.pep.*
- 11: /cgn2_6/ptodata/2/1aa/5K.COMB.pep.*
- 12: /cgn2_6/ptodata/2/1aa/5L.COMB.pep.*
- 13: /cgn2_6/ptodata/2/1aa/5M.COMB.pep.*
- 14: /cgn2_6/ptodata/2/1aa/5N.COMB.pep.*
- 15: /cgn2_6/ptodata/2/1aa/5O.COMB.pep.*
- 16: /cgn2_6/ptodata/2/1aa/5P.COMB.pep.*
- 17: /cgn2_6/ptodata/2/1aa/5Q.COMB.pep.*
- 18: /cgn2_6/ptodata/2/1aa/5R.COMB.pep.*
- 19: /cgn2_6/ptodata/2/1aa/5S.COMB.pep.*
- 20: /cgn2_6/ptodata/2/1aa/5T.COMB.pep.*
- 21: /cgn2_6/ptodata/2/1aa/5U.COMB.pep.*
- 22: /cgn2_6/ptodata/2/1aa/5V.COMB.pep.*
- 23: /cgn2_6/ptodata/2/1aa/5W.COMB.pep.*
- 24: /cgn2_6/ptodata/2/1aa/5X.COMB.pep.*
- 25: /cgn2_6/ptodata/2/1aa/5Y.COMB.pep.*
- 26: /cgn2_6/ptodata/2/1aa/5Z.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1179	100.0	218	1	US-08-470-837-26
2	1179	100.0	218	4	US-08-868-452-26
3	1179	100.0	231	4	US-08-129-722A-2
4	1179	100.0	232	4	US-08-377-874A-6
5	1179	100.0	232	5	PCT-US94-09700-6
6	1179	100.0	272	4	US-08-910-820-4
7	1179	100.0	272	4	US-08-910-820-6
8	1179	100.0	282	4	US-08-910-820-3
9	1179	100.0	282	4	US-08-910-820-5
10	1179	100.0	307	2	US-08-216-894-6
11	1179	100.0	307	4	US-09-115-746-6
12	1179	100.0	331	4	US-08-395-507-1
13	1179	100.0	352	1	US-08-216-894-4
14	1179	100.0	354	2	US-08-115-746-4
15	1179	100.0	354	4	US-08-395-507-2
16	1179	100.0	362	1	US-08-377-874A-11
17	1179	100.0	367	4	US-08-377-874A-11
18	1179	100.0	397	5	PCT-US94-09700-11
19	1179	100.0	412	4	US-09-366-009-34
20	1179	100.0	422	4	US-09-217-228-7
21	1179	100.0	426	3	US-08-737-248-4
22	1179	100.0	435	5	PCT-US93-04439-1
23	1179	100.0	439	4	US-08-506-2968-67
24	1179	100.0	442	4	US-08-506-2968-70
25	1179	100.0	443	4	US-08-506-2968-76
26	1179	100.0	447	4	US-08-506-2968-73
27	1179	100.0	472	2	US-08-216-894-10

28	1179	100.0	472	4	US-09-115-746-10	Sequence 10, Appl
29	1179	100.0	514	4	US-08-974-549A-605	Sequence 605, App
30	1179	100.0	515	4	US-08-974-549A-604	Sequence 604, App
31	1179	100.0	517	4	US-08-974-549A-606	Sequence 606, App
32	1179	100.0	530	4	US-08-974-549A-603	Sequence 603, App
33	1179	100.0	536	3	US-08-974-180-20	Sequence 20, Appl
34	1179	100.0	538	4	US-08-974-549A-602	Sequence 602, App
35	1179	100.0	542	4	US-08-506-2968-69	Sequence 69, Appl
36	1179	100.0	545	4	US-08-506-2968-75	Sequence 75, Appl
37	1179	100.0	547	4	US-08-506-2968-72	Sequence 72, Appl
38	1179	100.0	559	4	US-08-506-2968-66	Sequence 66, Appl
39	1179	100.0	564	2	US-08-216-894-2	Sequence 2, Appl
40	1179	100.0	564	4	US-09-115-746-2	Sequence 2, Appl
41	1179	100.0	579	2	US-08-864-224-11	Sequence 11, Appl
42	1179	100.0	632	4	US-08-506-2968-74	Sequence 74, Appl
43	1179	100.0	635	4	US-08-506-2968-71	Sequence 71, Appl
44	1179	100.0	643	2	US-08-216-894-8	Sequence 8, Appl
45	1179	100.0	643	4	US-09-115-746-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-08-470-837-26
Sequence 26, Application US/08470837
Patent No. 5800811

GENERAL INFORMATION:

APPLICANT: Nimm, Marcel E.
APPLICANT: Tuan, Tai-Lan
APPLICANT: Wu, Lingtao
APPLICANT: Cheung, David T.
TITLE OF INVENTION: Transforming Growth Factor B Fusion
TITLE OF INVENTION: and
TITLE OF INVENTION: Their Use in Wound Healing
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Merchant & Gould
STREET: 1150 Santa Monica Boulevard, Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025-3395

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,837
FILING DATE:
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Sharp, Janice A.
REGISTRATION NUMBER: 34,051
REFERENCE/DOCKET NUMBER: 30630-1US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 310-445-1140
TELEFAX: 310-445-9031

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:
LENGTH: 218 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-470-837-26

Query Match 100.0%; Score 1179; DB 1; Length 218;
Best Local Similarity 100.0%; Pred. No. 2.3e-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSPILGYWKIKGLVQPTRLLELYEYERHLYERDEGDKMKNKRFELGLEPNTLPYYID 60

```

Db 1 MSPIGLTWKIKGLVQPTRLLEYLEEKEEHLERDEGDKWRNKKFELGLEFPNLPYYID 60
    |||
Qy 61 GGVKLTQSMALIRYIADKHNMLGCGCPKERAISMLGAVLDIRYGVSRIVSKDFETLKV 120
    |||
Db 61 GGVKLTQSMALIRYIADKHNMLGCGCPKERAISMLGAVLDIRYGVSRIVSKDFETLKV 120
    |||
Qy 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPCDAPFKLVCFK 180
    |||
Db 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPCDAPFKLVCFK 180
    |||
Qy 181 KRIFAIPQIDKYLKSSKTYIAMPLOGWQATFGGSDHPK 218
    |||
Db 181 KRIFAIPQIDKYLKSSKTYIAMPLOGWQATFGGSDHPK 218
    |||

```

```

RESULT 2
US-08-868-452-26
; Sequence 26, Application US/08868452C
; Patent No. 6352972
; GENERAL INFORMATION:
; APPLICANT: Marcel E. Nimmi
; APPLICANT: Frederick L. Hall
; APPLICANT: Lingtao Wu
; APPLICANT: Bo Han
; APPLICANT: Edwin Shors
; TITLE OF INVENTION: BONE MORPHOGENETIC PROTEINS AND THEIR
; FILE REFERENCE: 17972-11
; CURRENT APPLICATION NUMBER: US/08/868,452C
; CURRENT FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 26
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-08-868-452-26

```

```

Query Match 100.0%; Score 1179; DB 4; Length 218;
Best Local Similarity 100.0%; Pred. No. 2.3e-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSPIGLTWKIKGLVQPTRLLEYLEEKEEHLERDEGDKWRNKKFELGLEFPNLPYYID 60
    |||
Db 1 MSPIGLTWKIKGLVQPTRLLEYLEEKEEHLERDEGDKWRNKKFELGLEFPNLPYYID 60
    |||
Qy 61 GGVKLTQSMALIRYIADKHNMLGCGCPKERAISMLGAVLDIRYGVSRIVSKDFETLKV 120
    |||
Db 61 GGVKLTQSMALIRYIADKHNMLGCGCPKERAISMLGAVLDIRYGVSRIVSKDFETLKV 120
    |||
Qy 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPCDAPFKLVCFK 180
    |||
Db 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPCDAPFKLVCFK 180
    |||
Qy 181 KRIFAIPQIDKYLKSSKTYIAMPLOGWQATFGGSDHPK 218
    |||
Db 181 KRIFAIPQIDKYLKSSKTYIAMPLOGWQATFGGSDHPK 218
    |||

```

```

RESULT 3
US-08-129-722A-2
; Sequence 2, Application US/08129722A
; Patent No. 6303369
; GENERAL INFORMATION:
; APPLICANT: Spana, Carl
; APPLICANT: Fargnoli, Joseph
; APPLICANT: Bolen, Joseph B.
; TITLE OF INVENTION: PROTEIN EXPRESSION SYSTEM
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burton Rodney
; STREET: P. O. Box 4000

```

```

; CITY: Princeton
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 08543-4000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/129,722A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaul, Timothy J.
; REGISTRATION NUMBER: 33,111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 252-5901
; TELEFAX: (609) 252-4526
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 231 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-129-722A-2

```

```

Query Match 100.0%; Score 1179; DB 4; Length 231;
Best Local Similarity 100.0%; Pred. No. 2.5e-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSPIGLTWKIKGLVQPTRLLEYLEEKEEHLERDEGDKWRNKKFELGLEFPNLPYYID 60
    |||
Db 1 MSPIGLTWKIKGLVQPTRLLEYLEEKEEHLERDEGDKWRNKKFELGLEFPNLPYYID 60
    |||
Qy 61 GGVKLTQSMALIRYIADKHNMLGCGCPKERAISMLGAVLDIRYGVSRIVSKDFETLKV 120
    |||
Db 61 GGVKLTQSMALIRYIADKHNMLGCGCPKERAISMLGAVLDIRYGVSRIVSKDFETLKV 120
    |||
Qy 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPCDAPFKLVCFK 180
    |||
Db 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYMDPCDAPFKLVCFK 180
    |||
Qy 181 KRIFAIPQIDKYLKSSKTYIAMPLOGWQATFGGSDHPK 218
    |||
Db 181 KRIFAIPQIDKYLKSSKTYIAMPLOGWQATFGGSDHPK 218
    |||

```

```

RESULT 4
US-08-327-874A-6
; Sequence 6, Application US/08327874A
; Patent No. 6372249
; GENERAL INFORMATION:
; APPLICANT: BAYLOR COLLEGE OF MEDICINE
; APPLICANT: SMITH, JAMES R.
; APPLICANT: DRUTZ, DAVID J.
; APPLICANT: WILSON, DEBORAH R.
; APPLICANT: ZUMSTEIN, LOUIS A.
; TITLE OF INVENTION: SENESCENT CELL DERIVED INHIBITORS OF
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROGERS & WELLS
; STREET: 200 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10166
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/327,874A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/808,523
FILING DATE: 16-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/970,462
FILING DATE: 02-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/113,372
FILING DATE: 30-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,564
FILING DATE: 17-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/203,535
FILING DATE: 25-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/229,420
FILING DATE: 15-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/274,535
FILING DATE: 13-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT US94/09700
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/327,874
FILING DATE: 24-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: NORTON, GERARD P.
REGISTRATION NUMBER: 36,621
REFERENCE/DOCKET NUMBER: 3634-8-CIP10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 878-3148
TELEFAX: (212) 878-8375
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 232 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Schistosoma japonicum
IMMEDIATE SOURCE:
CLONE: GST
US-08-327-874A-6

Query Match 100.0%; Score 1179; DB 4; Length 232;
Best Local Similarity 100.0%; Pred. No. 2,66-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGWKIKGLVPTLLLEYLEEKYEHLERDEGDKWNNKFFELGLEFPNLPYYID 60
DB 1 MSPILGWKIKGLVPTLLLEYLEEKYEHLERDEGDKWNNKFFELGLEFPNLPYYID 60
QY 61 GDVKTLSMAITRYADKHNMLGGCPKERAETISMLEGAVLDIRYGSRIAYSKDEFTLKV 120
DB 61 GDVKTLSMAITRYADKHNMLGGCPKERAETISMLEGAVLDIRYGSRIAYSKDEFTLKV 120
QY 121 DFLSLPEMLKMFEDRLCHKTLYLNDGHTHPDFNLXDALDVLVLYNDPKLDAFPLVYCK 180
DB 121 DFLSLPEMLKMFEDRLCHKTLYLNDGHTHPDFNLXDALDVLVLYNDPKLDAFPLVYCK 180
QY 181 KRIEATPOIDKYLKSSKYIAMPLOGMOTFFGGDHPK 218
DB 181 KRIEATPOIDKYLKSSKYIAMPLOGMOTFFGGDHPK 218

RESULT 5
PCT-US94-09700-6
Sequence 6, Application PC/TUS9409700
GENERAL INFORMATION:
APPLICANT: BAYLOR COLLEGE OF MEDICINE
APPLICANT: SMITH, JAMES R.
TITLE OF INVENTION: SENESCENT CELL DERIVED INHIBITORS OF
TITLE OF INVENTION: DNA SYNTHESIS
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOWREY & SIMON
STREET: 1299 PENNSYLVANIA AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09700
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/808,523
FILING DATE: 16-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/970,462
FILING DATE: 02-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/113,372
FILING DATE: 30-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,564
FILING DATE: 17-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/203,535
FILING DATE: 25-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/229,420
FILING DATE: 15-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/274,535
FILING DATE: 13-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: AUERBACH, JEFFREY I.
REGISTRATION NUMBER: 32,680
REFERENCE/DOCKET NUMBER: 225-102-CIP7-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 383-7451
TELEFAX: (202) 383-6610
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 232 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Schistosoma japonicum
IMMEDIATE SOURCE:
CLONE: GST
PCT-US94-09700-6

Query Match 100.0%; Score 1179; DB 5; Length 232;
Best Local Similarity 100.0%; Pred. No. 2,66-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGWKIKGLVPTLLLEYLEEKYEHLERDEGDKWNNKFFELGLEFPNLPYYID 60
DB 1 MSPILGWKIKGLVPTLLLEYLEEKYEHLERDEGDKWNNKFFELGLEFPNLPYYID 60

QY 61 GGVKLTOSMAIRIYIADKHNMLGGCPKERAEISMLEGAVLDIRGVSRISAKDEFTLV 120
DB 61 GGVKLTOSMAIRIYIADKHNMLGGCPKERAEISMLEGAVLDIRGVSRISAKDEFTLV 120
QY 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYLMDPCLDAFPKLVCFK 180
DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYLMDPCLDAFPKLVCFK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGWQATFGGDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGWQATFGGDHPK 218

RESULT 6

US-08-910-820-4
Sequence 4, Application US/08910820
Patent No. 6258579
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
APPLICANT: Zhu, Hengyi
APPLICANT: Barbosa, Miguel
APPLICANT: Li, Gfan
APPLICANT: Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
TITLE OF INVENTION: COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-910-820-4

Query Match 100.0%; Score 1179; DB 4; Length 272;

Best Local Similarity 100.0%; Pred. No. 3.2e-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHEHYERDEGDKMRKKFELGLEFPNLPYYID 60
DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHEHYERDEGDKMRKKFELGLEFPNLPYYID 60
QY 61 GGVKLTOSMAIRIYIADKHNMLGGCPKERAEISMLEGAVLDIRGVSRISAKDEFTLV 120
DB 61 GGVKLTOSMAIRIYIADKHNMLGGCPKERAEISMLEGAVLDIRGVSRISAKDEFTLV 120
QY 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYLMDPCLDAFPKLVCFK 180
DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYLMDPCLDAFPKLVCFK 180

QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGWQATFGGDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGWQATFGGDHPK 218

RESULT 7

US-08-910-820-6
Sequence 6, Application US/08910820
Patent No. 6258579
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
APPLICANT: Zhu, Hengyi
APPLICANT: Barbosa, Miguel
APPLICANT: Li, Gfan
APPLICANT: Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
TITLE OF INVENTION: COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-910-820-6

Query Match 100.0%; Score 1179; DB 4; Length 272;

Best Local Similarity 100.0%; Pred. No. 3.2e-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHEHYERDEGDKMRKKFELGLEFPNLPYYID 60
DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHEHYERDEGDKMRKKFELGLEFPNLPYYID 60
QY 61 GGVKLTOSMAIRIYIADKHNMLGGCPKERAEISMLEGAVLDIRGVSRISAKDEFTLV 120
DB 61 GGVKLTOSMAIRIYIADKHNMLGGCPKERAEISMLEGAVLDIRGVSRISAKDEFTLV 120
QY 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYLMDPCLDAFPKLVCFK 180
DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDMLYDALDVLYLMDPCLDAFPKLVCFK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGWQATFGGDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGWQATFGGDHPK 218

RESULT 8

US-08-910-820-3

Sequence 3, Application US/08910820
Patent No. 6258579
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
APPLICANT: Zhu, Hengyi
APPLICANT: Barbosa, Miguel
APPLICANT: Li, Gian
APPLICANT: Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
TITLE OF INVENTION: COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-910-820-3

Query Match 100.0%; Score 1179; DB 4; Length 282;
Best Local Similarity 100.0%; Pred. No. 3.3e-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 MSPILGYWKIKGLVOPTRLLLEKYEHEHLYERDEGDKWRKKFELGLEFNNLPYYID 60
DB 1 MSPILGYWKIKGLVOPTRLLLEKYEHEHLYERDEGDKWRKKFELGLEFNNLPYYID 60
QY 61 GDVKLTQSMALIRYIADKHNMLGGCPKERAIEISMLEGAVLDIRYGSRIAYSKDFETLVY 120
DB 61 GDVKLTQSMALIRYIADKHNMLGGCPKERAIEISMLEGAVLDIRYGSRIAYSKDFETLVY 120
QY 121 DFLSKLPENLKMFEEDRLCHRTYLNAGDHVTHPDMLYDALDVLYMDPMCLDAFPKLVCR 180
DB 121 DFLSKLPENLKMFEEDRLCHRTYLNAGDHVTHPDMLYDALDVLYMDPMCLDAFPKLVCR 180
QY 121 DFLSKLPENLKMFEEDRLCHRTYLNAGDHVTHPDMLYDALDVLYMDPMCLDAFPKLVCR 180
DB 121 DFLSKLPENLKMFEEDRLCHRTYLNAGDHVTHPDMLYDALDVLYMDPMCLDAFPKLVCR 180
QY 181 KRLEAIPOIDKYLKSSKYIAMPLOGQATFGGSDHPPK 218
DB 181 KRLEAIPOIDKYLKSSKYIAMPLOGQATFGGSDHPPK 218

RESULT 9
US-08-910-820-5
Sequence 5, Application US/08910820
Patent No. 6258579
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
APPLICANT: Zhu, Hengyi
APPLICANT: Barbosa, Miguel
APPLICANT: Li, Gian
APPLICANT: Murray, Brian W.

TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
TITLE OF INVENTION: COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-910-820-5

Query Match 100.0%; Score 1179; DB 4; Length 282;
Best Local Similarity 100.0%; Pred. No. 3.3e-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVOPTRLLLEKYEHEHLYERDEGDKWRKKFELGLEFNNLPYYID 60
DB 1 MSPILGYWKIKGLVOPTRLLLEKYEHEHLYERDEGDKWRKKFELGLEFNNLPYYID 60
QY 61 GDVKLTQSMALIRYIADKHNMLGGCPKERAIEISMLEGAVLDIRYGSRIAYSKDFETLVY 120
DB 61 GDVKLTQSMALIRYIADKHNMLGGCPKERAIEISMLEGAVLDIRYGSRIAYSKDFETLVY 120
QY 121 DFLSKLPENLKMFEEDRLCHRTYLNAGDHVTHPDMLYDALDVLYMDPMCLDAFPKLVCR 180
DB 121 DFLSKLPENLKMFEEDRLCHRTYLNAGDHVTHPDMLYDALDVLYMDPMCLDAFPKLVCR 180
QY 181 KRLEAIPOIDKYLKSSKYIAMPLOGQATFGGSDHPPK 218
DB 181 KRLEAIPOIDKYLKSSKYIAMPLOGQATFGGSDHPPK 218

RESULT 10
US-08-216-894-6
Sequence 6, Application US/08216894
Patent No. 5876734
GENERAL INFORMATION:
APPLICANT: Kirchhoff, Louis V.
APPLICANT: Otsu, Keiko
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
TITLE OF INVENTION: WITH TRYPANOSOMA CRUZI
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/216,894
FILING DATE: 24-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 85326/102/DRLO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 307 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-216-894-6

Query Match 100.0%; Score 1179; DB 2; Length 307;
Best Local Similarity 100.0%; Pred. No. 3.8e-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMNKKFELGLEFPNLPYYID 60
OY 61 GGVKLTQSMALIRYIADKHNMLGGCPKERAETSMLEGAVALDIRYGVSRIVASKDEFTLV 120
DB 61 GGVKLTQSMALIRYIADKHNMLGGCPKERAETSMLEGAVALDIRYGVSRIVASKDEFTLV 120
OY 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHYTHPDPMLYDALDVLYNDPMCLDAFPKLVCFK 180
DB 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHYTHPDPMLYDALDVLYNDPMCLDAFPKLVCFK 180
OY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFGGDDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFGGDDHPK 218

RESULT 11
US-09-115-746-6
Sequence 6, Application US/09115746
Patent No. 6228601
GENERAL INFORMATION:
APPLICANT: Kirchhoff, Louis V.
APPLICANT: Otsu, Keiko
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
TITLE OF INVENTION: WITH TRYPAPOSOMA CRUZI
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/115,746
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/216,894
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 85326/102/DRLO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 307 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-115-746-6

Query Match 100.0%; Score 1179; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 3.8e-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMNKKFELGLEFPNLPYYID 60
OY 61 GGVKLTQSMALIRYIADKHNMLGGCPKERAETSMLEGAVALDIRYGVSRIVASKDEFTLV 120
DB 61 GGVKLTQSMALIRYIADKHNMLGGCPKERAETSMLEGAVALDIRYGVSRIVASKDEFTLV 120
OY 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHYTHPDPMLYDALDVLYNDPMCLDAFPKLVCFK 180
DB 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHYTHPDPMLYDALDVLYNDPMCLDAFPKLVCFK 180
OY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFGGDDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFGGDDHPK 218

RESULT 12
US-09-217-228-6
Sequence 6, Application US/09217228
Patent No. 6323178
GENERAL INFORMATION:
APPLICANT: Butler, Jon P.
APPLICANT: Hale, John E.
APPLICANT: Heath Jr., William F.
APPLICANT: Schoner, Brigitte E.
APPLICANT: Helman, Mark L.
APPLICANT: Becker, Gerald W.
APPLICANT: Varshavsky, Alexander D.
TITLE OF INVENTION: Beta-11pocrofin and Uses Thereof
FILE REFERENCE: X-12139
CURRENT APPLICATION NUMBER: US/09/217,228
CURRENT FILING DATE: 1998-12-21
NUMBER OF SEQ ID NOS: 35
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 6
LENGTH: 331
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: GST fusion
OTHER INFORMATION: partner
US-09-217-228-6

Query Match 100.0%; Score 1179; DB 4; Length 331;
Best Local Similarity 100.0%; Pred. No. 4.2e-117;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMNKKFELGLEFPNLPYYID 60
OY 61 GGVKLTQSMALIRYIADKHNMLGGCPKERAETSMLEGAVALDIRYGVSRIVASKDEFTLV 120
DB 61 GGVKLTQSMALIRYIADKHNMLGGCPKERAETSMLEGAVALDIRYGVSRIVASKDEFTLV 120

QY 121 DFLSKLPEMLKMFEDRLCHRTYLNQDHTHPDMLYDALDVLVYMDPCLDAFPKLYCFK 180
DB 121 DFLSKLPEMLKMFEDRLCHRTYLNQDHTHPDMLYDALDVLVYMDPCLDAFPKLYCFK 180
QY 181 KRIEAIPOIDKYLKSSKTYIAMPLOGWQATFGGDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKTYIAMPLOGWQATFGGDHPPK 218

RESULT 13

US-08-395-507-1
Sequence 1, Application US/08395507
Patent No. 5578456
GENERAL INFORMATION:
APPLICANT: Fujimura, Katsuya
APPLICANT: Ueno, Eiichi
APPLICANT: Fujii, No. 5578456yuk1
APPLICANT: Okada, Masahisa
TITLE OF INVENTION: Anti-treponema Pallidum Antibody
TITLE OF INVENTION: Immunoassay
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pollock, Vande Sande & Priddy
STREET: 1990 M Street, NW, Suite 800
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/395,507
FILING DATE: 27-FEB-1995
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP Hei-6-54672
FILING DATE: 28-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: Amerdick, Burton A.
REGISTRATION NUMBER: 24,852
REFERENCE/DOCKET NUMBER: 1581/007
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 331-7111
TELEFAX: (202) 293-6229
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 352 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: rabbit
FEATURE:
NAME/KEY: Protein
LOCATION: 1..352
OTHER INFORMATION: /note= "G15 Antigen"

Query Match 100.0%; Score 1179; DB 1; Length 352;

Best Local Similarity 100.0%; Pred. No. 4.5e-117; Mismatches 0; Indels 0; Gaps 0;

Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGWKIKGLVQPTRLLEYLEEKYEHLERDGDGDKWRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGWKIKGLVQPTRLLEYLEEKYEHLERDGDGDKWRNKKFELGLEFPNLPYYID 60
QY 61 GDVKTOSMAIIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGVSRIVSKDFETLVK 120

DB 61 GDVKTOSMAIIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGVSRIVSKDFETLVK 120
QY 121 DFLSKLPEMLKMFEDRLCHRTYLNQDHTHPDMLYDALDVLVYMDPCLDAFPKLYCFK 180
DB 121 DFLSKLPEMLKMFEDRLCHRTYLNQDHTHPDMLYDALDVLVYMDPCLDAFPKLYCFK 180
QY 181 KRIEAIPOIDKYLKSSKTYIAMPLOGWQATFGGDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKTYIAMPLOGWQATFGGDHPPK 218

RESULT 14

US-08-216-894-4
Sequence 4, Application US/08216894
Patent No. 5876734
GENERAL INFORMATION:
APPLICANT: Kirchoff, Louis V.
APPLICANT: Otsu, Keiko
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/216,894
FILING DATE: 24-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 85326/102/DRLO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 354 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-216-894-4

Query Match 100.0%; Score 1179; DB 2; Length 354;

Best Local Similarity 100.0%; Pred. No. 4.6e-117; Mismatches 0; Indels 0; Gaps 0;

Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGWKIKGLVQPTRLLEYLEEKYEHLERDGDGDKWRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGWKIKGLVQPTRLLEYLEEKYEHLERDGDGDKWRNKKFELGLEFPNLPYYID 60
QY 61 GDVKTOSMAIIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGVSRIVSKDFETLVK 120
DB 61 GDVKTOSMAIIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGVSRIVSKDFETLVK 120
QY 121 DFLSKLPEMLKMFEDRLCHRTYLNQDHTHPDMLYDALDVLVYMDPCLDAFPKLYCFK 180
DB 121 DFLSKLPEMLKMFEDRLCHRTYLNQDHTHPDMLYDALDVLVYMDPCLDAFPKLYCFK 180
QY 181 KRIEAIPOIDKYLKSSKTYIAMPLOGWQATFGGDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKTYIAMPLOGWQATFGGDHPPK 218

RESULT 15
US-09-115-746-4
Sequence 4, Application US/09115746
Patent No. 6228601
GENERAL INFORMATION:
APPLICANT: Kirchhoff, Louis V.
APPLICANT: Otsu, Keiko
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/115,746
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/216,894
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 85326/102/DRLO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 354 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-115-746-4

Query Match 100.0%; Score 1179; DB 4; Length 354;
Best Local Similarity 100.0%; Pred. No. 4: 6e-117; Indels 0; Gaps 0;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSPILGYWKINGLVOPTRILEYLEEKYEHLIERDEGDKWRNKKFELGLEFPNLPYIID 60
|||
Db 1 MSPILGYWKINGLVOPTRILEYLEEKYEHLIERDEGDKWRNKKFELGLEFPNLPYIID 60
QY 61 GDVKLTQSMALIRIYADKHNHNLGGCPKERAEISMLEGAVLDIRYGVSRIRAYSKDEFILKV 120
|||
Db 61 GDVKLTQSMALIRIYADKHNHNLGGCPKERAEISMLEGAVLDIRYGVSRIRAYSKDEFILKV 120
QY 121 DFLSKLPMLKMFEDRLCHKTYLLNGDHTHPDMLYDALDVLYNDPMLCLDAFPKLVCFK 180
|||
Db 121 DFLSKLPMLKMFEDRLCHKTYLLNGDHTHPDMLYDALDVLYNDPMLCLDAFPKLVCFK 180
QY 181 KRIEAIPOIDKYLKSKSYIAMPLOGWQATFGGCHHPK 218
|||
Db 181 KRIEAIPOIDKYLKSKSYIAMPLOGWQATFGGCHHPK 218

Search completed: May 20, 2003, 11:40:56
Job time : 8.02587 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:37:59 ; Search time 34.7335 Seconds

(Without alignments)
622.443 Million cell updates/sec

Title: US-10-081-408-4

Perfect score: 1179

Sequence: 1 MSPILGYWKIKGLVQPTRL.....IAMPLOGWATFGGDDHPK 218

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEM_PUB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/FCI_NEM_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEM_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/PCRTS_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEM_PUB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEM_PUB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEM_PUB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1179	100.0	218	9	US-10-081-408-4
2	1179	100.0	239	10	US-09-823-153-7
3	1179	100.0	245	9	US-10-267-311-23
4	1179	100.0	272	9	US-09-844-988-4
5	1179	100.0	272	9	US-09-844-988-6
6	1179	100.0	272	10	US-09-844-908-4
7	1179	100.0	272	10	US-09-844-908-6
8	1179	100.0	282	9	US-09-844-988-3
9	1179	100.0	282	10	US-09-844-988-5
10	1179	100.0	282	10	US-09-844-908-3
11	1179	100.0	282	10	US-09-844-908-5
12	1179	100.0	298	9	US-09-910-600-25
13	1179	100.0	324	9	US-10-267-311-25
14	1179	100.0	331	9	US-09-824-438-6
15	1179	100.0	348	9	US-09-910-600-22
16	1179	100.0	348	9	US-09-910-600-23
17	1179	100.0	348	9	US-09-910-600-24
18	1179	100.0	348	9	US-09-910-600-25
19	1179	100.0	354	10	US-09-823-153-8

20	1179	100.0	364	9	US-09-788-268-14	Sequence 14, Appl
21	1179	100.0	394	10	US-09-990-578-4	Sequence 4, Appl
22	1179	100.0	401	9	US-10-190-866A-1	Sequence 1, Appl
23	1179	100.0	412	9	US-09-775-964-34	Sequence 34, Appl
24	1179	100.0	422	9	US-09-824-438-7	Sequence 7, Appl
25	1179	100.0	440	9	US-09-950-634-4	Sequence 4, Appl
26	1176	99.7	649	9	US-10-174-784-9	Sequence 9, Appl
27	1149	97.5	218	9	US-10-081-408-5	Sequence 5, Appl
28	1144	97.0	988	9	US-10-081-408-20	Sequence 20, Appl
29	472.5	40.1	229	9	US-10-102-806-554	Sequence 554, App
30	187.5	15.9	222	9	US-09-784-738-4	Sequence 4, Appl
31	187.5	15.9	230	10	US-09-925-301-1318	Sequence 1318, Ap
32	171.5	14.5	222	9	US-09-784-739-3	Sequence 3, Appl
33	165.5	14.0	222	9	US-09-784-739-5	Sequence 5, Appl
34	164.5	14.0	203	9	US-09-847-208-48	Sequence 48, Appl
35	152.5	12.9	222	9	US-09-784-738-1	Sequence 1, Appl
36	99	8.4	240	9	US-09-964-899-23	Sequence 23, Appl
37	99	8.4	379	9	US-10-253-007-46	Sequence 46, Appl
38	97	8.2	233	10	US-09-765-213A-2	Sequence 2, Appl
39	94.5	8.0	412	9	US-09-214-592-31	Sequence 31, Appl
40	92	7.8	233	10	US-09-765-213A-4	Sequence 4, Appl
41	89	7.5	233	10	US-09-765-213A-6	Sequence 6, Appl
42	87.5	7.4	359	9	US-09-738-626-4895	Sequence 4895, Ap
43	83	7.0	1668	10	US-09-815-242-5654	Sequence 5654, Ap
44	83	7.0	2397	10	US-09-815-242-12265	Sequence 12265, A
45	81.5	6.9	241	9	US-09-854-133-204	Sequence 204, App

ALIGNMENTS

RESULT 1
US-10-081-408-4
; Sequence 4, Application US/10081408
; Patent No. US20020160482A1
; GENERAL INFORMATION:
; APPLICANT: Abirams, n, Lars
; TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
; FILE REFERENCE: 13425-053001
; CURRENT APPLICATION NUMBER: US/10/081,408
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: SE 0100625-3
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 60/272,247
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FASTED for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Schistosoma japonicum
US-10-081-408-4

Query Match 100.0%; Score 1179; DB 9; Length 218;
Best Local Similarity 100.0%; Pred. No. 2e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSPILGYWKIKGLVQPTRLLEYLEKYEHEHLYERDEGCKMKKKFELGLEPPNLPYYID	60
DB	1	MSPILGYWKIKGLVQPTRLLEYLEKYEHEHLYERDEGCKMKKKFELGLEPPNLPYYID	60
QY	61	GVVLTLSAAIRYIADKNNMLGGCPKEARETSMLEGAVALDIRYGVSRIRAYSKDPETLKV	120
DB	61	GVVLTLSAAIRYIADKNNMLGGCPKEARETSMLEGAVALDIRYGVSRIRAYSKDPETLKV	120
QY	121	DELTKLPEMLKMFEDRLCKRTYLANDVHTHPFMYDALDVLYLXDPKCLDAFPKLYCCK	180
DB	121	DELTKLPEMLKMFEDRLCKRTYLANDVHTHPFMYDALDVLYLXDPKCLDAFPKLYCCK	180
QY	181	KRIEAIPOIDKYLKSSKTIYAMPLOGWATFGGDDHPK 218	
DB	181	KRIEAIPOIDKYLKSSKTIYAMPLOGWATFGGDDHPK 218	

RESULT 2
US-09-823-153-7
; Sequence 7, Application US/09823153
; Patent No. US20020025540A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; APPLICANT: Roberts, Susan
; APPLICANT: Pak, Roger
; APPLICANT: Lewis, Martin
; APPLICANT: Smith, David
; APPLICANT: Hendrick, Joseph
; APPLICANT: Vinitsky, Alexander
; TITLE OF INVENTION: ISOLATION OF FUNCTIONALLY ACTIVE GAMMA-SECRETASE PROTEIN COMPLEX
; TITLE OF INVENTION: METHODS FOR DETECTION OF ACTIVITY AND INHIBITORS THEREOF
; FILE REFERENCE: DD004
; CURRENT APPLICATION NUMBER: US/09/823.153
; CURRENT FILING DATE: 2001-07-02
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 7
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Description of Artificial Sequence: PSI PEPTIDE
US-09-823-153-7

Query Match 100.0%; Score 1179; DB 10; Length 239;
Best Local Similarity 100.0%; Pred. No. 2,2e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGVKIKIGLVOPRLLLEYLEEKEEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGVKIKIGLVOPRLLLEYLEEKEEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
QY 61 GDVKLTOSMAIRYIADKHNMLGGCPKERAISMLEGAVLDIRGVSRIVASKDFETLKV 120
DB 61 GDVKLTOSMAIRYIADKHNMLGGCPKERAISMLEGAVLDIRGVSRIVASKDFETLKV 120
QY 121 DFLSKLPMLKMFEDRLCHKTYLNGDHYTHDPFMYLDALDVLYMDPCLDAFPKLVCFK 180
DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHYTHDPFMYLDALDVLYMDPCLDAFPKLVCFK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOTFGGCDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOTFGGCDHPK 218

RESULT 3
US-10-267-311-23
; Sequence 23, Application US/10267311
; Publication No. US20030050469A1
; GENERAL INFORMATION:
; APPLICANT: Siegel, Marvin
; APPLICANT: Chu, N. Randall
; APPLICANT: Mizeu, Lee A.
; TITLE OF INVENTION: INDUCTION OF A TH1-LIKE RESPONSE IN VITRO
; FILE REFERENCE: 12071/002001
; CURRENT APPLICATION NUMBER: US/10/267.311
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US/09/613.303
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: US 60/143.757
; PRIOR FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 23
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Artificial Sequence

FEATURE:
; OTHER INFORMATION: fusion sequence
US-10-267-311-23

Query Match 100.0%; Score 1179; DB 9; Length 245;
Best Local Similarity 100.0%; Pred. No. 2,3e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGVKIKIGLVOPRLLLEYLEEKEEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGVKIKIGLVOPRLLLEYLEEKEEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
QY 61 GDVKLTOSMAIRYIADKHNMLGGCPKERAISMLEGAVLDIRGVSRIVASKDFETLKV 120
DB 61 GDVKLTOSMAIRYIADKHNMLGGCPKERAISMLEGAVLDIRGVSRIVASKDFETLKV 120
QY 121 DFLSKLPMLKMFEDRLCHKTYLNGDHYTHDPFMYLDALDVLYMDPCLDAFPKLVCFK 180
DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHYTHDPFMYLDALDVLYMDPCLDAFPKLVCFK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOTFGGCDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOTFGGCDHPK 218

RESULT 4
US-09-844-988-4
; Sequence 4, Application US/09844988
; Patent No. US20020158764A1
; GENERAL INFORMATION:
; APPLICANT: Mercurio, Frank
; APPLICANT: Zhu, Hengyi
; APPLICANT: Barbosa, Miguel
; APPLICANT: Li, Gian
; APPLICANT: Murray, Brian W.
; TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
; TITLE OF INVENTION: COMPLEX AND METHODS OF USE THEREFOR
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESS: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/844.988
; FILING DATE: 26-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/910.820
; FILING DATE: 1997-08-13
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-844-988-4

Query Match 100.0%; Score 1179; DB 9; Length 272;
Best Local Similarity 100.0%; Pred. No. 2.7e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPIIGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKWRNKKFELGLEFPNLPYYID 60
DB 1 MSPIIGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKWRNKKFELGLEFPNLPYYID 60
QY 61 GVKYLTQSMALIRYIADKHNMLGGCPKRAEISMLGAVLDIRYGSRIAYSKDFTLKY 120
DB 61 GVKYLTQSMALIRYIADKHNMLGGCPKRAEISMLGAVLDIRYGSRIAYSKDFTLKY 120
QY 121 DFLSKPEMLKMEDELCKHTYINGDHVTHPDMLYDALDVLYXMDPCIDAPKLYCCK 180
DB 121 DFLSKPEMLKMEDELCKHTYINGDHVTHPDMLYDALDVLYXMDPCIDAPKLYCCK 180
QY 181 KRTEAIPOIDKYLKSSKSIYAMPLOGQATFGGSDHPPK 218
DB 181 KRTEAIPOIDKYLKSSKSIYAMPLOGQATFGGSDHPPK 218

RESULT 5

US-09-844-988-6
Sequence 6, Application US/09844988
Patent No. US20020158764A1

GENERAL INFORMATION:

APPLICANT: Mercurio, Frank
Zhu, Hengyi
Barbosa, Miguel
Li, Gian
Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/844,988
FILING DATE: 26-Apr-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/910,820
FILING DATE: 1997-08-13

ATTORNEY/AGENT INFORMATION:

NAME: Makl, David J.
REGISTRATION NUMBER: 31,392

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-09-844-988-6
Query Match 100.0%; Score 1179; DB 9; Length 272;
Best Local Similarity 100.0%; Pred. No. 2.7e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPIIGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKWRNKKFELGLEFPNLPYYID 60
DB 1 MSPIIGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKWRNKKFELGLEFPNLPYYID 60
QY 61 GVKYLTQSMALIRYIADKHNMLGGCPKRAEISMLGAVLDIRYGSRIAYSKDFTLKY 120
DB 61 GVKYLTQSMALIRYIADKHNMLGGCPKRAEISMLGAVLDIRYGSRIAYSKDFTLKY 120
QY 121 DFLSKPEMLKMEDELCKHTYINGDHVTHPDMLYDALDVLYXMDPCIDAPKLYCCK 180
DB 121 DFLSKPEMLKMEDELCKHTYINGDHVTHPDMLYDALDVLYXMDPCIDAPKLYCCK 180
QY 181 KRTEAIPOIDKYLKSSKSIYAMPLOGQATFGGSDHPPK 218
DB 181 KRTEAIPOIDKYLKSSKSIYAMPLOGQATFGGSDHPPK 218

RESULT 6

US-09-844-908-4
Sequence 4, Application US/09844908
Patent No. US20020151021A1

GENERAL INFORMATION:

APPLICANT: Mercurio, Frank
Zhu, Hengyi
Barbosa, Miguel
Li, Gian
Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/844,908
FILING DATE: 27-Apr-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997

ATTORNEY/AGENT INFORMATION:

NAME: Makl, David J.
REGISTRATION NUMBER: 31,392

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-09-844-908-4
Query Match 100.0%; Score 1179; DB 10; Length 272;
Best Local Similarity 100.0%; Pred. No. 2.7e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPIIGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKWRNKKFELGLEFPNLPYYID 60
DB 1 MSPIIGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKWRNKKFELGLEFPNLPYYID 60

QY 61 GGVKLTQSAIIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGVSRAYSKDFTLV 120
DB 61 GGVKLTQSAIIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGVSRAYSKDFTLV 120
QY 121 DFLSKLPKMFEDRLCKHTYLNQDHTHPDFMLYDALDVLYMDPKLDAFPKLVCFK 180
DB 121 DFLSKLPKMFEDRLCKHTYLNQDHTHPDFMLYDALDVLYMDPKLDAFPKLVCFK 180
QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMQATFGGDDHPK 218
DB 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMQATFGGDDHPK 218

RESULT 7

US-09-844-908-6
Sequence 6, Application US/09844908
Patent No. US20020151021A1
GENERAL INFORMATION:

APPLICANT: Mercurio, Frank

Zhu, Hengyi
Barbosa, Miguel
Li, Gfan

Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR

NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA

ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,908
FILING DATE: 27-Apr-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS: <unknown>
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-844-908-6

Query Match 100.0%; Score 1179; DB 10; Length 272;
Best Local Similarity 100.0%; Pred. No. 2.7e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGWYKIKGLVPTRLLEYLEEKYEHLHYERDEGDKWNKKFELGLEPNNLPYYID 60
DB 1 MSPILGWYKIKGLVPTRLLEYLEEKYEHLHYERDEGDKWNKKFELGLEPNNLPYYID 60
QY 61 GGVKLTQSAIIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGVSRAYSKDFTLV 120
DB 61 GGVKLTQSAIIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGVSRAYSKDFTLV 120

QY 121 DFLSKLPKMFEDRLCKHTYLNQDHTHPDFMLYDALDVLYMDPKLDAFPKLVCFK 180
DB 121 DFLSKLPKMFEDRLCKHTYLNQDHTHPDFMLYDALDVLYMDPKLDAFPKLVCFK 180
QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMQATFGGDDHPK 218
DB 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMQATFGGDDHPK 218

RESULT 8

US-09-844-988-3
Sequence 3, Application US/09844988
Patent No. US20020158764A1
GENERAL INFORMATION:

APPLICANT: Mercurio, Frank

Zhu, Hengyi
Barbosa, Miguel
Li, Gfan

Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR

NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA

ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,988
FILING DATE: 26-Apr-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/910,820
FILING DATE: 1997-08-13
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 amino acids
TYPE: amino acid
STRANDEDNESS: <unknown>
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-844-988-3

Query Match 100.0%; Score 1179; DB 9; Length 282;
Best Local Similarity 100.0%; Pred. No. 2.8e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGWYKIKGLVPTRLLEYLEEKYEHLHYERDEGDKWNKKFELGLEPNNLPYYID 60
DB 1 MSPILGWYKIKGLVPTRLLEYLEEKYEHLHYERDEGDKWNKKFELGLEPNNLPYYID 60
QY 61 GGVKLTQSAIIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGVSRAYSKDFTLV 120
DB 61 GGVKLTQSAIIRYIADKHNMLGCGPKERAISMLEGAVLDIRYGVSRAYSKDFTLV 120
QY 121 DFLSKLPKMFEDRLCKHTYLNQDHTHPDFMLYDALDVLYMDPKLDAFPKLVCFK 180
DB 121 DFLSKLPKMFEDRLCKHTYLNQDHTHPDFMLYDALDVLYMDPKLDAFPKLVCFK 180

QY 181 KRLEAIPOIDKYLKSSKYIAMPLOGWQATFGGDDHPPK 218
 Db 181 KRLEAIPOIDKYLKSSKYIAMPLOGWQATFGGDDHPPK 218

RESULT 9

US-09-844-988-5

Sequence 5, Application US/09844988

Patent No. US20020158764A1

GENERAL INFORMATION:

APPLICANT: Mercurio, Frank

Zhu, Hengyi

Barbosa, Miguel

Li, Gian

Murray, Brian W.

TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE

COMPLEX AND METHODS OF USE THEREFOR

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/844,988

FILING DATE: 26-Apr-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/910,820

FILING DATE: 1997-08-13

ATTORNEY/AGENT INFORMATION:

NAME: Makl, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 860098.413C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 282 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-09-844-988-5

Query Match 100.0%; Score 1179; DB 9; Length 282;

Best Local Similarity 100.0%; Pred. No. 2.8e-111;

Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGWMKIKGLVOPTRLLLEYLEEKYEHEHLYERDEGDKWRKKKTELGLFEPNLPYYID 60
 Db 1 MSPILGWMKIKGLVOPTRLLLEYLEEKYEHEHLYERDEGDKWRKKKTELGLFEPNLPYYID 60
 QY 61 GGVKLTQSAIIRYIADKNHMLGGCCKERAIEISMLGAVLDIRYGSRIAYSDFETLKY 120
 Db 61 GGVKLTQSAIIRYIADKNHMLGGCCKERAIEISMLGAVLDIRYGSRIAYSDFETLKY 120
 QY 121 DFLSKLPEMLKMFEDRLCKRTYLNQDHTVHPDFMLYDALDVLYMDPMLDAFPLKVCER 180
 Db 121 DFLSKLPEMLKMFEDRLCKRTYLNQDHTVHPDFMLYDALDVLYMDPMLDAFPLKVCER 180
 QY 181 KRLEAIPOIDKYLKSSKYIAMPLOGWQATFGGDDHPPK 218
 Db 181 KRLEAIPOIDKYLKSSKYIAMPLOGWQATFGGDDHPPK 218

RESULT 10
 US-09-844-908-3

Sequence 3, Application US/09844908

Patent No. US20020151021A1

GENERAL INFORMATION:

APPLICANT: Mercurio, Frank

Zhu, Hengyi

Barbosa, Miguel

Li, Gian

Murray, Brian W.

TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE

COMPLEX AND METHODS OF USE THEREFOR

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/844,908

FILING DATE: 27-Apr-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/910,820

FILING DATE: 12-AUG-1997

ATTORNEY/AGENT INFORMATION:

NAME: Makl, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 860098.413C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 282 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 3:

US-09-844-908-3

Query Match 100.0%; Score 1179; DB 10; Length 282;

Best Local Similarity 100.0%; Pred. No. 2.8e-111;

Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSPILGWMKIKGLVOPTRLLLEYLEEKYEHEHLYERDEGDKWRKKKTELGLFEPNLPYYID 60
 Db 1 MSPILGWMKIKGLVOPTRLLLEYLEEKYEHEHLYERDEGDKWRKKKTELGLFEPNLPYYID 60
 QY 61 GGVKLTQSAIIRYIADKNHMLGGCCKERAIEISMLGAVLDIRYGSRIAYSDFETLKY 120
 Db 61 GGVKLTQSAIIRYIADKNHMLGGCCKERAIEISMLGAVLDIRYGSRIAYSDFETLKY 120
 QY 121 DFLSKLPEMLKMFEDRLCKRTYLNQDHTVHPDFMLYDALDVLYMDPMLDAFPLKVCER 180
 Db 121 DFLSKLPEMLKMFEDRLCKRTYLNQDHTVHPDFMLYDALDVLYMDPMLDAFPLKVCER 180
 QY 181 KRLEAIPOIDKYLKSSKYIAMPLOGWQATFGGDDHPPK 218
 Db 181 KRLEAIPOIDKYLKSSKYIAMPLOGWQATFGGDDHPPK 218

RESULT 11

US-09-844-908-5

Sequence 5, Application US/09844908

```

; Patent No. US20020151021A1
; GENERAL INFORMATION:
; APPLICANT: Mercurio, Frank
;              Zhu, Hengyi
;              Barbosa, Miguel
;              Li, Gian
;              Murray, Brian W.
; TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
;                   COMPLEX AND METHODS OF USE THEREFOR
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/844,908
; FILING DATE: 27-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,820
; FILING DATE: 12-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 860098.413C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 282 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-844-908-5
Query Match          100.0%; Score 1179; DB 10; Length 282;
Best Local Similarity 100.0%; Pred. No. 2.8e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSPILGWKIRIGLVOPTRLLLEYLEKEYEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGWKIRIGLVOPTRLLLEYLEKEYEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
QY 61 GDVKLTQSMATIRIYADKHNMLGCGPKERAEISMLEGAVLDIRGVSRIVASKDFETLKV 120
DB 61 GDVKLTQSMATIRIYADKHNMLGCGPKERAEISMLEGAVLDIRGVSRIVASKDFETLKV 120
QY 121 DFLSKLPEMLKMFEDRLCHKTYTLNGDHTHPDMLYDALDVLYMDPCLDAPFKLVCFK 180
DB 121 DFLSKLPEMLKMFEDRLCHKTYTLNGDHTHPDMLYDALDVLYMDPCLDAPFKLVCFK 180
QY 181 KRIEAIPOIDKYLKSSKXIAMPLOGWQATFGGCDHPK 218
DB 181 KRIEAIPOIDKYLKSSKXIAMPLOGWQATFGGCDHPK 218
; RESULT 12
US-09-910-600-26
; Sequence 26, Application US/09910600
; Publication No. US20030036631A1
; GENERAL INFORMATION:
; APPLICANT: Longphre, Malinda
; APPLICANT: Chang, Han

```

```

; APPLICANT: Whitney, Gena
; TITLE OF INVENTION: NOVEL SIGLECS AND USES THEREOF
; FILE REFERENCE: D0003NP
; CURRENT APPLICATION NUMBER: US/09/910,600
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 60/220,139
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 26
; LENGTH: 298
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: L3cyto-Y641
US-09-910-600-26
Query Match          100.0%; Score 1179; DB 9; Length 298;
Best Local Similarity 100.0%; Pred. No. 3e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSPILGWKIRIGLVOPTRLLLEYLEKEYEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGWKIRIGLVOPTRLLLEYLEKEYEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
QY 61 GDVKLTQSMATIRIYADKHNMLGCGPKERAEISMLEGAVLDIRGVSRIVASKDFETLKV 120
DB 61 GDVKLTQSMATIRIYADKHNMLGCGPKERAEISMLEGAVLDIRGVSRIVASKDFETLKV 120
QY 121 DFLSKLPEMLKMFEDRLCHKTYTLNGDHTHPDMLYDALDVLYMDPCLDAPFKLVCFK 180
DB 121 DFLSKLPEMLKMFEDRLCHKTYTLNGDHTHPDMLYDALDVLYMDPCLDAPFKLVCFK 180
QY 181 KRIEAIPOIDKYLKSSKXIAMPLOGWQATFGGCDHPK 218
DB 181 KRIEAIPOIDKYLKSSKXIAMPLOGWQATFGGCDHPK 218
; RESULT 13
US-10-267-311-25
; Sequence 25, Application US/10267311
; Publication No. US20030050469A1
; GENERAL INFORMATION:
; APPLICANT: Siegel, Marvin
; APPLICANT: Chu, N. Randall
; APPLICANT: Mitzen, Lee A.
; TITLE OF INVENTION: INDUCTION OF A TH1-LIKE RESPONSE IN VITRO
; FILE REFERENCE: 12071/002001
; CURRENT APPLICATION NUMBER: US/10/267,311
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US/09/613,303
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: US 60/143,757
; PRIOR FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion sequence
US-10-267-311-25
Query Match          100.0%; Score 1179; DB 9; Length 324;
Best Local Similarity 100.0%; Pred. No. 3.3e-111;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSPILGWKIRIGLVOPTRLLLEYLEKEYEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGWKIRIGLVOPTRLLLEYLEKEYEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60

```

Tue May 20 13:19:15 2003

us-10-081-408-4.rapb

Page 7

[illegible]

```

RESULT 14
US-09-824-438-6
Sequence 6, Application US/09824438
Publication NO. US20030073621A1
GENERAL INFORMATION:
APPLICANT: Butler, Jon P.
APPLICANT: Hale, John E.
APPLICANT: Heath Jr., William F.
APPLICANT: Schoner, Brigitte E.
APPLICANT: Helman, Mark L.
APPLICANT: Becker, Gerald W.
APPLICANT: Varshavsky, Alexander D.
TITLE OF INVENTION: Beta-lipotropin and Uses Thereof
FILE REFERENCE: X-12139
CURRENT APPLICATION NUMBER: US/09/824,438
CURRENT FILING DATE: 2001-10-29
NUMBER OF SEQ ID NOS: 35
SOFTWARE: PatentIn Ver. 2.0
Seq ID NO 6
LENGTH: 331
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: GST fusion
US-09-824-438-6

```

	Query Match	100.0%	Score 1179;	DB 9;	Length 331.
	Best Local Similarity	100.0%	Pred. No. 3 4e-111;		
	Matches 218; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;
Oy	1 MSPILGYNKINGLVOPTRLLLEYLEEKYEENLYHERDEGDKNWKKFELGLEFPNLPYYID	60			
Db	1 MSPILGYMKIKGLVOPTRLLLEYLEEKYEENLYHERDEGDCKWNKKFELGLEFPNPYYID	60			
Oy	61 GGVKLTSQMAIIRYADNHNMIGGPKREAEISMEGAVALDIRGVSRAYSKEDETLY	120			
Db	61 GGVKLTSQMAIIRYADNHNMIGGPKREAEISMEGAVALDIRGVSRAYSKEDETLY	120			
Oy	121 DFLSLPEMLKMFEDRLCHKTLYLNGDHYTHDPFMLYDALDVYLXNDPMDLAFAPLYCFK	180			
Db	121 DFLSLPEMLKMFEDRLCHKTLYLNGDHYTHDPFMLYDALDVYLXNDPMDLAFAPLYCFK	180			
Oy	181 KRIEAIPIQIDKYLKSSKIYAMPLOQMOTFFGGSDHPRK	218			
Db	181 KRIEAIPIQIDKYLKSSKIYAMPLOQMOTFFGGSDHPRK	218			

RESULT 15
US-09-910-600-22
Sequence 22, Application US/09910600
Publication NO. US20030036631A1
GENERAL INFORMATION:
APPLICANT: Longphre, Malinda
APPLICANT: Chang, Han
APPLICANT: Whitley, Gena
TITLE OF INVENTION: NOVEL SICLECS AND USES THEREOF
FILE REFERENCE: D0003ND
CURRENT APPLICATION NUMBER: US/09/910,600
CURRENT FILING DATE: 2001-07-20

? PRIOR APPLICATION NUMBER: 60/220,139
 ? PRIOR FILING DATE: 2000-07-21
 ? NUMBER OF SEQ ID NOS: 32
 ? SOFTWARE: PatentIn Ver. 2.0
 ? SEQ ID NO 22
 ? LENGTH: 348
 ? TYPE: PR1
 ? ORGNISM: Artificial Sequence
 ? FEATURE:
 ? OTHER INFORMATION: Description of Artificial Sequence: l3cyto-wt
 ? IS-09-910-600-22

	Query Match	100.0%;	Score 1179;	DB 9;	Length 348;
	Best Local Similarity	100.0%;	Pred. No. 3,7e-111;		
	Matches 218;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MSPILGIYKIKICLVOPTRILLEYLBEKYEYBEHLERDEGCKRNKKKFFELGELERNPLPYIID	60		
DB	1	MSPILGIYKIKICLVOPTRILLEYLBEKYEYBEHLERDEGCKRNKKKFFELGELERNPLPYIID	60		
QY	61	GVKRLTQSMAIRRYIADKHNMLGGCPKREAEISMLEGAVLDIRYGVSRVAYSKEDEETLKV	120		
DB	61	GVKRLTQSMAIRRYIADKHNMLGGCPKREAEISMLEGAVLDIRYGVSRVAYSKEDEETLKV	120		
QY	121	DFLSLKPEMLKMFEDRLCHKTYLNGDHYTHDPFMTLYDALDYVLKMDPMCLDAFPKLVCFK	180		
DB	121	DFLSLKPEMLKMFEDRLCHKTYLNGDHYTHDPFMTLYDALDYVLKMDPMCLDAFPKLVCFK	180		
QY	181	KRIEAIPOIDKYLKSSKSYIAMPLOGQOAFEGGDDHPK	218		
DB	181	KRIEAIPOIDKYLKSSKSYIAMPLOGQOAFEGGDDHPK	218		

Search completed: May 20, 2003, 12:02:21
Job time : 35.7335 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:09:54 ; Search time 8.41126 Seconds

(without alignments)
2491.578 Million cell updates/sec

Title: US-10-081-408-4

Perfect score: 1179

Sequence: 1 MSPILGKWKIKGLVQPTRL.....IAMPLOGQWATFGGDHPK 218

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR73:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1170	99.2	219	2	A26484 glutathione transt
2	990	84.0	218	2	A45556 glutathione S-tran
3	927	78.6	218	2	A45523 glutathione transt
4	654	55.5	209	2	A48388 glutathione S-tran
5	483.5	41.0	218	2	S33860 glutathione transt
6	483.5	41.0	218	2	A29794 glutathione transt
7	480.5	40.8	218	2	S32425 glutathione transt
8	480.5	40.8	218	2	A47486 glutathione transt
9	478.5	40.6	218	2	S65674 glutathione transt
10	474.5	40.2	218	2	A39375 glutathione transt
11	473.5	40.2	217	2	JX0095 glutathione transt
12	470.5	39.9	218	2	S13202 glutathione transt
13	469.5	39.8	218	2	A23732 glutathione transt
14	468.5	39.7	218	2	B34159 glutathione transt
15	467.5	39.7	218	2	S01719 glutathione transt
16	464.5	39.4	218	2	B29331 glutathione transt
17	464.5	39.4	218	2	A46048 glutathione transt
18	462.5	39.2	218	2	A29036 glutathione transt
19	456.5	38.7	218	1	XURTG4 glutathione transt
20	453.5	38.5	218	2	A46143 mu-class glutathio
21	453.5	38.5	218	2	B28946 glutathione transt
22	449.5	38.1	225	2	A35295 glutathione transt
23	429.5	36.4	230	2	S18464 glutathione transt
24	385.5	32.7	219	2	S50146 major allergen Dpl
25	257.5	21.6	142	2	S17462 glutathione transt
26	248.5	21.1	208	2	S41933 glutathione transt
27	244	20.7	210	2	S71958 glutathione transt
28	242	20.5	203	2	S13780 glutathione transt
29	240	20.4	210	1	A37378 glutathione transt

30	238	20.2	210	2	S71957 glutathione transt
31	237	20.1	210	1	XURTG4 glutathione transt
32	237	20.1	210	1	B55140 glutathione transt
33	236	20.0	210	2	I48112 glutathione S-tran
34	236	20.0	210	2	S71959 glutathione transt
35	235	19.9	210	2	JC6529 glutathione transt
36	233	19.8	209	1	A55140 glutathione transt
37	232	19.7	210	2	A41177 glutathione transt
38	230	19.5	210	2	A49180 glutathione transt
39	219.5	18.6	208	2	S03615 glutathione transt
40	214.5	18.2	223	1	A41031 glutathione transt
41	200.5	17.0	222	2	A26653 glutathione transt
42	198.5	16.8	223	1	A27848 glutathione transt
43	195.5	16.6	222	1	XURTG4 glutathione transt
44	193.5	16.4	221	2	A54858 glutathione transt
45	193.5	16.4	221	2	S24322 glutathione transt

ALIGNMENTS

```
RESULT 1
A26484
glutathione transferase (EC 2.5.1.18) - fluke (Schistosoma japonicum) (fragment)
C:Species: Schistosoma japonicum
C>Date: 25-Oct-1987 #sequence_revision 30-Sep-1989 #text_change 03-Feb-1994
C:Accession: A94139; A26484; A28315
R:Smith, D.B.; Davern, K.M.; Board, P.G.; Tiu, W.U.; Garcia, E.G.; Mitchell, G.F.
Proc. Natl. Acad. Sci. U.S.A. 83, 8703-8707, 1986
A:Title: Mr 26,000 antigen of Schistosoma japonicum recognized by resistant WEHI 1:
A:Reference number: A94139; MUID:87041520; PMID:3095841
A:Accession: A94139
A:Molecule type: mRNA
A:Residues: 1-219 <SMID>
R:Smith, D.B.; Davern, K.M.; Board, P.G.; Tiu, W.U.; Garcia, E.G.; Mitchell, G.F.
Proc. Natl. Acad. Sci. U.S.A. 84, 6541, 1987
A:Reference number: A94181
A:Contents: annotation; revision to residues 210-219
A:Keywords: glutathione transferase
C:Keywords: transferase

Query Match          99.2%: Score 1170; DB 2; Length 219;
Best Local Similarity 100.0%: Pred. No. 9.6e-95;
Matches 216; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PILGKWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMRKKFELGEPNLPYIIDG 62
    |||||||
Db 4 PILGKWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMRKKFELGEPNLPYIIDG 63
    |||||||

QY 63 VKLTOSMAIIRYIADKNHMLGCGCPKRAEISMLGCAVIDIRYGSRIAYSDFETLKADF 122
    |||||||
Db 64 VKLTOSMAIIRYIADKNHMLGCGCPKRAEISMLGCAVIDIRYGSRIAYSDFETLKADF 123
    |||||||

QY 123 LSKPEMLKMEDERLCKHTYVNGDHVHPDMLDALDVYVYMPMCADAPKVCFFKR 182
    |||||||
Db 124 LSKPEMLKMEDERLCKHTYVNGDHVHPDMLDALDVYVYMPMCADAPKVCFFKR 183
    |||||||

QY 183 IEAIPQIDKYLLKSSKRYIAMPLOGQWATFGGDHPK 218
    |||||||
Db 184 IEAIPQIDKYLLKSSKRYIAMPLOGQWATFGGDHPK 219
    |||||||

RESULT 2
A45556
glutathione S-transferase - fluke (Schistosoma mansoni)
C:Species: Schistosoma mansoni
C>Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 03-May-1996
C:Accession: A45556
R:Wright, M.D.; Harrison, R.A.; Meider, A.M.; Newport, G.R.; Mitchell, G.F.
Mol. Biochem. Parasitol. 49, 177-179, 1991
A:Title: Another 26-kilodalton glutathione S-transferase of Schistosoma mansoni.
A:Reference number: A45556; MUID:92131046; PMID:1775156
A:Accession: A45556
```

A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-218 <NR>
 A>Note: sequence extracted from NCBI backbone (NCBI:77814)
 C:Superfamily: glutathione transferase

Query Match 84.0%; Score 990; DB 2; Length 218;
 Best Local Similarity 82.6%; Pred. No. 4.8e-79;
 Matches 180; Conservative 19; Mismatches 19; Indels 0; Gaps 0;

OY 1 MSPILGYWKINGLVOPTRLLLEYLEEKYEENLYERDEGDKMKNKKFELGEPNLPYYID 60
 DB 1 MAPKLGWKIKGLVOPTRLLLEYLGERYERLYDNRDGVNNEKFKGLDFPNLPYYID 60
 OY 61 GGVKLTQSMALIRYIADKHNMLGGCPKERAETSMLEGAVIDIRYGSRIATYSGDEPTLVY 120
 DB 61 GGVKLTQSMALIRYIADKHNMLGGCPKERAETSMLEGAVIDIRYGSRIATYSGDEPTLVY 120
 OY 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTVHPDFMLYDALDVLYMDPMLDAFPKLVCFK 180
 DB 121 DFLNQLPGLKMFEDRLSHNTYLNNDKVTYHPDFMLYDALDVLYMDPMLDAFPKLVCFK 180
 OY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMATFEGGDHPK 218
 DB 181 ORIENLPRIKYNLNSRIYIKMPLQGSATFEGGDAPK 218

RESULT 3
 A:5523
 glutathione transferase (EC 2.5.1.18) - fluke (Schistosoma mansoni)
 C:Species: Schistosoma mansoni
 C:Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 18-Jun-1999
 C:Accession: A45523; A4941
 R:Trouten, F.; Kleny, M.P.; Vervaeke, C.; Torpier, G.; Pierce, R.J.; Balloul, J.M.; Schol. Blochem. Parasitol. 41, 35-44, 1990
 A:Title: Molecular cloning and tissue distribution of a 26-kilodalton Schistosoma mansoni
 A:Reference number: A45523; MUID:90348716; PMID:2385266
 A:Accession: A45523
 A:Molecule type: mRNA
 A:Status: preliminary
 A:Residues: 1-218 <TR>
 A:Cross-references: GB:M41106; NID:g161004; PID:AAA29888.1; PID:g552241
 R:Henkle, K.J.; Davern, K.M.; Wright, M.D.; Ramos, A.J.; Mitchell, G.F.
 Mol. Biochem. Parasitol. 40, 23-34, 1990
 A:Title: Comparison of the cloned genes of the 26- and 28-kilodalton glutathione S-trans
 A:Reference number: A44941; MUID:90271935; PMID:1693415
 A:Accession: A44941
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 8-218 <HE>
 A:Cross-references: GB:M26914; GB:M26913; NID:g161006; PID:AAA29889.1; PID:g161007
 C:Superfamily: glutathione transferase
 C:Keywords: transferase

Query Match 78.6%; Score 927; DB 2; Length 218;
 Best Local Similarity 79.8%; Pred. No. 1.5e-73;
 Matches 174; Conservative 15; Mismatches 29; Indels 0; Gaps 0;

OY 1 MSPILGYWKIKLVOPTRLLLEYLEEKYEENLYERDEGDKMKNKKFELGEPNLPYYID 60
 DB 1 MAPKLGWKIKGLVOPTRLLLEYLGERYERLYDNRDGVNNEKFKGLDFPNLPYYID 60
 OY 61 GGVKLTQSMALIRYIADKHNMLGGCPKERAETSMLEGAVIDIRYGSRIATYSGDEPTLVY 120
 DB 61 GGVKLTQSMALIRYIADKHNMLGGCPKERAETSMLEGAVIDIRYGSRIATYSGDEPTLVY 120
 OY 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTVHPDFMLYDALDVLYMDPMLDAFPKLVCFK 180
 DB 121 DFLNQLPGLKMFEDRLSHNTYLNNDKVTYHPDFMLYDALDVLYMDPMLDAFPKLVCFK 180
 OY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMATFEGGDHPK 218
 DB 181 KCIEDLPRIKYNLNSRIYIKMPLQGSATFEGGDAPK 218

RESULT 4

A:4838
 glutathione S-transferase - liver fluke (fragment)
 C:Species: Fasciola hepatica (liver fluke)
 C:Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 03-May-1996
 C:Accession: A4838
 R:Muio, A.; Rodriguez-Molina, J.R.; Hillyer, G.V.
 Am. J. Trop. Med. Hyg. 48, 457-463, 1993

A:Title: Sequence analysis of a Fasciola hepatica glutathione S-transferase cDNA clone
 A:Reference number: A4838; MUID:93228188; PMID:7682383
 A:Accession: A4838

A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-209 <MR>
 A:Experimental source: Puerto Rico
 A>Note: sequence extracted from NCBI backbone (NCBI:129138, NCBI:129139)
 C:Superfamily: glutathione transferase

Query Match 55.5%; Score 654; DB 2; Length 209;
 Best Local Similarity 58.2%; Pred. No. 9.6e-50;
 Matches 121; Conservative 31; Mismatches 56; Indels 0; Gaps 0;

OY 10 IKGVLVOPTRLLLEYLEEKYEENLYERDEGDKMKNKKFELGEPNLPYYIDGDKLTQSM 69
 DB 1 IRGLQVAVLLLEYLGERYERLYDNRDGVNNEKFKGLDFPNLPYYIDCKLTQSM 60
 OY 70 AIIRYIADKHNMLGGCPKERAETSMLEGAVIDIRYGSRIATYSGDEPTLVY 129
 DB 61 AIIRYIADKHNMLGGCPKERAETSMLEGAVIDIRYGSRIATYSGDEPTLVY 120
 OY 130 LKMFEDRLCHKTYLNGDHTVHPDFMLYDALDVLYMDPMLDAFPKLVCFKRIEAIPOI 189
 DB 121 LKMFEDRLCHKTYLNGDHTVHPDFMLYDALDVLYMDPMLDAFPKLVCFKRIEAIPOI 180
 OY 190 DYLKSSKIYAMPLOGMATFEGGDHP 217
 DB 181 KAVMESEKRIKYNLNSRIYIKMPLQGSATFEGGDAP 208

RESULT 5

A:53860
 glutathione transferase (EC 2.5.1.18) mu1 - mouse
 N:Alternative names: glutathione transferase major chain; glutathione transferase mu-8.
 C:Species: Mus musculus (house mouse)
 C:Date: 08-Dec-1993 #sequence_revision 07-Oct-1994 #text_change 18-Jun-1999
 C:Accession: S33860; A34159; A28946; A20831; I24735; S16933; S71312
 R:Reinhart, J.; Pearson, W.R.
 Arch. Biochem. Biophys. 303, 383-393, 1993
 A:Title: The structure of two murine class-mu glutathione transferase genes coordinat
 A:Reference number: S33860; MUID:93290350; PMID:8512323
 A:Accession: S33860

A:Molecule type: DNA
 A:Residues: 1-218 <REI>
 A:Cross-references: GB:I13448
 A>Note: the authors translated the codon GAG for residue 49 as Gly and GAG for residu
 R:Townsend, A.J.; Goldsmith, M.E.; Pickett, C.B.; Cowan, K.H.
 J. Biol. Chem. 264, 21582-21590, 1989

A:Title: Isolation, characterization, and expression in Escherichia coli of two murin
 A:Reference number: A34159; MUID:90094327; PMID:2689439
 A:Accession: A34159

A:Molecule type: mRNA
 A:Residues: 1-218 <TON>
 A:Cross-references: GB:J04632; NID:g193547; PID:AAA37705.1; PID:g309260
 R:Pearson, W.R.; Reinhart, J.; Sisk, S.C.; Anderson, K.S.; Adler, P.N.
 J. Biol. Chem. 263, 13324-13332, 1988

A:Title: Tissue-specific induction of murine glutathione transferase mRNAs by butylat
 A:Reference number: A92668; MUID:88330838; PMID:3417659
 A:Contents: Clone PGT 875

A:Accession: A28946
 A:Molecule type: mRNA
 A:Residues: 2-218 <PEA>

A:Cross-references: GB:J03952; NID:9193687; PIDN:AAA37747.1; PID:9309278
 R:Pearson, W.R.; Windle, J.J.; Morrow, J.F.; Benson, A.M.; Talalay, P.
 J. Biol. Chem. 258, 2052-2062, 1983
 A:Title: Increased synthesis of glutathione S-transferases in response to anticarcinogen
 A:Reference number: A02411; MUID:83109018; PMID:6822548
 A:Accession: A20831
 A:Molecule type: protein
 A:Residues: 2-7, 'X', 8-13, 'X', 15-23, 'X', 25, 'XX', 28, 'N', 30-31, 'X', 33, 'X', 35-39, 'W', 41 <PEI
 R:Manerlyk, B.; Alin, P.; Guttenberg, C.; Jenson, H.; Tahir, M.K.; Warholm, M.; Jorvæ
 Proc. Natl. Acad. Sci. U.S.A. 82, 7202-7206, 1985
 A:Title: Identification of three classes of cytosolic glutathione transferase common to
 A:Reference number: A24735; MUID:86042634; PMID:3864155
 A:Accession: I24735
 A:Molecule type: protein
 A:Residues: 2-7, 'X', 8-13, 'X', 15-23, 'X', 25 <MAN>
 R:Hayes, J.D.; Kerr, L.A.; Peacock, S.D.; Cronshaw, A.D.; Mclellan, L.I.
 Biochem. J. 277, 501-512, 1991
 A:Title: Hepatic glutathione S-transferases in mice fed on a diet containing the anticar
 adent elution of the glutathione-S-transferase affinity matrix.
 A:Reference number: S16933; MUID:91315425; PMID:1859377
 A:Accession: S16933
 A:Molecule type: protein
 A:Residues: 2-36 <NAV>
 R:Fernandes, C.L.; Dong, J.H.; Roebuck, B.D.; Chisari, F.V.; Montali, J.A.; Schmidt Jr.,
 Arch. Biochem. Biophys. 331, 104-116, 1996
 A:Title: Elevations of hepatic quinone reductase, glutathione, and alpha- and mu-class g
 A:Reference number: S71311; MUID:96268466; PMID:8660889
 A:Accession: S71312
 A:Molecule type: protein
 A:Residues: 97-108 <FEK>
 C:Genetics:
 A:Gene: GSTM1
 A:Intons: 12/3; 38/1; 59/3; 87/1; 120/3; 152/3; 189/3
 C:Superfamily: glutathione transferase
 C:Keywords: dimer; transferase
 F:2-218/Product: glutathione transferase mul #status experimental <MAT>
 Query Match 41.0%; Score 483.5; DB 2; Length 218;
 Best Local Similarity 44.2%; Pred. No. 7.6e-35;
 Matches 92; Conservative 38; Mismatches 73; Indels 5; Gaps 1;
 QY 1 MSPILGYWKIKGVOPTRLLLEYLEEYEEHLERDEG-----DKRNKKEFELGLEPRL 55
 DB 1 MPMLITGMWNRGLTHPLRLLETTDSYDEKRTMGADPDRSOWINERFKGLDPLNL 60
 QY 56 PYIIDGVKLTQSMATIRYADKNHMGCGPKERAETISMEGAVLDIRYGVSRIVASKDF 115
 DB 61 PYLIDSHKRTQSNALIRYARKNHNLGTEEEERIRADIVENQVMOTRMOLIMCTVPPDF 120
 QY 116 ETLKVDPLSKLPKMLKMFEDRLCHKTYLNGDHTVHPDFMLDALDVVLYMDPMLDAEPK 175
 DB 121 EKQKPEFLKTIPEKMKLYSEFLGKRPFADGKTYVDFLAYDILDQYRMEPCIDAFPN 180
 QY 176 LVCFKKRIEAIPOIDKYIKSSKIYAMP 203
 DB 181 LRDFLAPEGLKRIISAMKSSRYIATPI 208
 RESULT 6
 A29794
 glutathione transferase (EC 2.5.1.18) class mu chain 3 - rat
 N:Alternate names: glutathione S-transferase Yb1; ligandin
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 28-Aug-1989 #sequence revision 28-Aug-1989 #text change 18-Jun-1999
 R:Accession: A29794; A25510; A24085; B61004; F24735; A33397; S27111; S17167; A26187
 R:Chang, C.; Saltzman, A.G.; Sorensen, N.S.; Hlipkka, R.A.; Liao, S.
 J. Biol. Chem. 262, 11901-11903, 1987
 A:Title: Identification of glutathione S-transferase Yb-1 mRNA as the androgen-repressed
 A:Reference number: A29794; MUID:87308179; PMID:3040722
 A:Accession: A29794
 A:Molecule type: mRNA
 A:Residues: 1-218 <CHA>
 A:Cross-references: GB:J02810; NID:9204514; PIDN:AAA1293.1; PID:9204515

R:Lat, H.C.J.; Grove, G.; Tu, C.P.D.
 Nucleic Acids Res. 14, 6101-6114, 1986
 A:Title: Cloning and sequence analysis of a cDNA for a rat liver glutathione S-tran
 A:Reference number: A25510; MUID:86312882; PMID:2875437
 A:Accession: A25510
 A:Molecule type: mRNA
 A:Residues: 1-218 <LAT>
 A:Cross-references: GB:X04229; NID:956337; PIDN:CAA27811.1; PID:956338
 R:Ping, G.J.F.; Lu, A.Y.H.; Pickett, C.B.
 J. Biol. Chem. 260, 13268-13271, 1985
 A:Reference number: A24085; MUID:86033768; PMID:3840477
 A:Accession: A24085
 A:Molecule type: mRNA
 A:Residues: 1-198, 'NC', 201-218 <DIN>
 A:Cross-references: GB:M11719; NID:9204502; PIDN:AAA1287.1; PID:9204503
 A:Experimental source: clone pGTR/C44
 R:Chang, L.H.; Hsieh, J.C.; Chen, W.L.; Tam, M.F.
 Electrophoresis 11, 589-593, 1990
 A:Title: Identification of rat liver glutathione S-transferase Yb subunits by parti
 ical isoelectric focusing gel.
 A:Reference number: A61004; MUID:91031411; PMID:2226415
 A:Accession: B61004
 A:Molecule type: protein
 A:Residues: 2-26 <CH2>
 R:Manerlyk, B.; Alin, P.; Guttenberg, C.; Jenson, H.; Tahir, M.K.; Warholm, M.; J
 Proc. Natl. Acad. Sci. U.S.A. 82, 7202-7206, 1985
 A:Title: Identification of three classes of cytosolic glutathione transferase commc
 A:Reference number: A24735; MUID:86042634; PMID:3864155
 A:Accession: F24735
 A:Molecule type: protein
 A:Residues: 2-20 <MAN>
 R:Hsieh, J.C.; Liu, L.F.; Chen, W.L.; Tam, M.F.
 Biochem. Biophys. Res. Commun. 162, 1147-1154, 1989
 A:Title: Expression of Yb-1 glutathione S-transferase using a baculovirus expressic
 A:Reference number: A33397; MUID:89350924; PMID:2669745
 A:Accession: A33397
 A:Molecule type: protein
 A:Residues: 2-21; 212-218 <HS1>
 R:Katuz, R.M.; Bono, B.; Colman, R.F.
 Arch. Biochem. Biophys. 298, 667-677, 1992
 A:Title: Identification of Tyr(115) labeled by S-(4-Bromo-2,3-dioxobutyl)glutathior
 A:Reference number: S27111; MUID:93037509; PMID:1416995
 A:Accession: S27111
 A:Molecule type: protein
 A:Residues: 2-20; 83-86, 'X', 88-96; 109-115, 'X', 117-122 <KAT>
 R:Hsieh, J.C.; Huang, S.C.; Chen, W.L.; Lai, Y.C.; Tam, M.F.
 Biochem. J. 278, 293-297, 1991
 A:Title: Cysteine-86 is not needed for the enzymic activity of glutathione S-transf
 A:Reference number: S17167; MUID:91354218; PMID:1683338
 A:Accession: S17167
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 2-218 <HS2>
 C:Superfamily: glutathione transferase
 C:Keywords: dimer; transferase
 Query Match 41.0%; Score 483.5; DB 2; Length 218;
 Best Local Similarity 44.2%; Pred. No. 7.6e-35;
 Matches 92; Conservative 38; Mismatches 73; Indels 5; Gaps 1;
 QY 1 MSPILGYWKIKGVOPTRLLLEYLEEYEEHLERDEG-----DKRNKKEFELGLEPRL 55
 DB 1 MPMLITGMWNRGLTHPLRLLETTDSYDEKRTMGADPDRSOWINERFKGLDPLNL 60
 QY 56 PYIIDGVKLTQSMATIRYADKNHMGCGPKERAETISMEGAVLDIRYGVSRIVASKDF 115
 DB 61 PYLIDSHKRTQSNALIRYARKNHNLGTEEEERIRADIVENQVMOTRMOLIMCTVPPDF 120
 QY 116 ETLKVDPLSKLPKMLKMFEDRLCHKTYLNGDHTVHPDFMLDALDVVLYMDPMLDAEPK 175
 DB 121 EKQKPEFLKTIPEKMKLYSEFLGKRPFADGKTYVDFLAYDILDQYRMEPCIDAFPN 180
 QY 176 LVCFKKRIEAIPOIDKYIKSSKIYAMP 203

Db 181 LKDFLARFEGLEKISAYMKSSRYLSTPI 208

RESULT 7

S32425

glutathione transferase (EC 2.5.1.18) class mu, GSTM4 (version 2) - human

N:Alternate names: glutathione transferase mu4

C:Species: Homo sapiens (man)

C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 18-Jun-1999

C:Accession: S32425; S29337

R:Zhong, S.; Spurr, N.K.; Hayes, J.D.; Wolf, C.R.

Biochem. J. 291, 41-50, 1993

A:Title: Deduced amino acid sequence, gene structure and chromosomal location of a novel

A:Reference number: S32424; MUID:93228631; PMID:8471052

A:Accession: S32425

A:Molecule type: DNA

A:Residues: 1-218 <2HO>

A:Cross-references: EMBL:X68677; NID:g31934; PIDN:CAA4837.1; PID:g31935

A:Note: the authors translated the codon AAG for residue 182 as Arg and CCA for residue

C:Genetics:

A:Gene: GDB:GSTM4

A:Cross-references: GDB:134191; OMIM:138333

A:Map position: 1p13.3-1p13.3

A:Introns: 12/3; 38/1; 59/3; 87/1; 120/3; 152/3; 189/3

C:Superfamily: glutathione transferase

C:Keywords: dimer; transferase

Query Match 40.8%; Score 480.5; DB 2; Length 218;

Best Local Similarity 44.0%; Pred. No. 1.4e-34;

Matches 95; Conservative 37; Mismatches 79; Indels 5; Gaps 1;

1 MSPILGYWKIKINGVQPTRLLEVEEKEEHLRYERD-----EGKMKRKKELEPPNL 55
1 MPMLGFWDIRGLAHARILLETYDSSYEKKYMGAPDYDRSQWLNKKRLGLDPPNL 60
56 PYIDGDKVLTQSMALIRYADKHNMLGCKPERAEISMLEGAVLDIRYGSRTAISKDF 115
61 PYLDGAKKITQSNAILCYIAKRNHLCGETEEKIRVDILENQAMDVSNOQLARCYSPDF 120
116 ETLKVDLSKLPEMLKMFEDRLCHKTYLNGDVTHTPDMFLYDALDLYLMDPMLCLDAFPK 175
121 EKLPEVLEELPTJMQHRSQFLGKRPWFVGDKITFVDFLAYVDLHRIFFEPNCLDAFPN 180
176 LVCFKKRIEALPQIDKYLKSSKITYAMPLOGQWATFG 211
181 LKDFISREFGLEKISAYMKSSRFLPKPLYTRVAVWG 216

RESULT 8

A47486

glutathione transferase (EC 2.5.1.18) class mu, GSTM4 (version 1) - human

C:Species: Homo sapiens (man)

C>Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 18-Jun-1999

C:Accession: A47486; B47486; S36782; I37438; S45685

R:Johnson, K.E.; Johnson, K.J.; Rifkenbery, D.; Henner, W.D.

J. Biol. Chem. 268, 16958-16965, 1993

A:Title: Isolation and analysis of the gene and cDNA for a human mu class glutathione S-

A:Reference number: A47486; MUID:93352467; PMID:8349586

A:Accession: A47486

A:Molecule type: DNA

A:Residues: 1-218 <COM>

A:Cross-references: GB:M96233

A:Accession: B47486

A:Molecule type: mRNA

A:Residues: 1-218 <COM>

A:Cross-references: GB:M96234; NID:g306818; PIDN:AAA57347.1; PID:g306819

A:Experimental source: Hela cells

R:Rosa, V.L.; Board, P.G.

Biochem. J. 294, 373-380, 1993

A:Title: Molecular cloning and heterologous expression of an alternatively spliced human

A:Reference number: S36782; MUID:93384505; PMID:8373352

A:Accession: S36782

A:Molecule type: mRNA

A:Residues: 1-218 <ROS>

A:Cross-references: EMBL:M99422

A:Experimental source: testis

R:Comstock, K.E.; Windersten, M.; Hao, X.Y.; Henner, W.D.; Mannervik, B.

Arch. Biochem. Biophys. 311, 487-495, 1994

A:Title: A comparison of the enzymatic and physicochemical properties of human glutat

A:Reference number: S45684; MUID:94263230; PMID:8203914

A:Contents: annotation

R:Taylor, J.B.; Oliver, J.; Sherrington, R.; Pemble, S.E.

Biochem. J. 274, 587-593, 1991

A:Title: Structure of human glutathione S-transferase class Mu genes.

A:Reference number: I37438; MUID:91174774; PMID:2006920

A:Accession: I37438

A:Molecule type: DNA

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Residues: 39-120 <RES>

A:Cross-references: EMBL:X68637; NID:g31936; PIDN:CAA0167.1; PID:g31937

C:Genetics:

A:Gene: GDB:GSTM4

A:Cross-references: GDB:134191; OMIM:138333

A:Map position: 1p13.3-1p13.3

A:Introns: 12/3; 38/1; 59/3; 87/1; 120/3; 152/3; 189/3

C:Superfamily: glutathione transferase

C:Keywords: alternative splicing; dimer; transferase

Query Match 40.8%; Score 480.5; DB 2; Length 218;

Best Local Similarity 44.0%; Pred. No. 1.4e-34;

Matches 95; Conservative 37; Mismatches 79; Indels 5; Gaps 1;

1 MSPILGYWKIKINGVQPTRLLEVEEKEEHLRYERD-----DKMKRKKELEPPNL 55
1 MPMLGFWDIRGLAHARILLETYDSSYEKKYMGAPDYDRSQWLNKKRLGLDPPNL 60
56 PYIDGDKVLTQSMALIRYADKHNMLGCKPERAEISMLEGAVLDIRYGSRTAISKDF 115
61 PYLDGAKKITQSNAILCYIAKRNHLCGETEEKIRVDILENQAMDVSNOQLARCYSPDF 120
116 ETLKVDLSKLPEMLKMFEDRLCHKTYLNGDVTHTPDMFLYDALDLYLMDPMLCLDAFPK 175
121 EKLPEVLEELPTJMQHRSQFLGKRPWFVGDKITFVDFLAYVDLHRIFFEPNCLDAFPN 180
176 LVCFKKRIEALPQIDKYLKSSKITYAMPLOGQWATFG 211
181 LKDFISREFGLEKISAYMKSSRFLPKPLYTRVAVWG 216

RESULT 9

S65674

glutathione transferase (EC 2.5.1.18) class mu chain 7.8 - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 18-Jun-1999

C:Accession: S65674; S30380

R:Lee, S.H.; Lee, Y.S.; Han, J.S.; Kim, Y.S.; Koh, J.K.

Arch. Biochem. Biophys. 318, 424-429, 1995

A:Title: Cloning and expression of a cDNA for Mu-Class Glutathione S-transferase from

A:Reference number: S65674; MUID:95251394; PMID:7733673

A:Accession: S65674

A:Molecule type: mRNA

A:Residues: 1-218 <LEP>

A:Cross-references: EMBL:L23766; NID:g388198; PIDN:AAA69665.1; PID:g388199

R:Primiano, T.; Novak, R.F.

Arch. Biochem. Biophys. 301, 404-410, 1993

A:Title: Purification and characterization of class mu glutathione S-transferase iso2

A:Reference number: S30380; MUID:93213177; PMID:8460949

A:Accession: S30380

A:Molecule type: Protein

A:Residues: 2-21 <PRI>

C:Superfamily: glutathione transferase

A:Experimental source: liver
 R:Seidegard, J.; Vorachek, W.R.; Pero, R.W.; Pearson, W.R.
 Proc. Natl. Acad. Sci. U.S.A. 85, 7293-7297, 1988
 A:Title: Hereditary differences in the expression of the human glutathione transferase A
 A:Reference number: A30770; MUID:89017184; PMID:3174634
 A:Contents: allele 1b (psi)
 A:Accession: A30770
 A:Molecule type: mRNA
 A:Residues: 1-172, 'N', 174-218 <SEI>
 A:Cross-references: EMBL:003817; NID:9183668; PIDN:AAA59203.1; PID:g306812
 A:Experimental source: liver
 R:Zhong, S.; Spurr, N.K.; Hayes, J.D.; Wolf, C.R.
 Biochem. J. 291, 41-50, 1993
 A:Title: Deduced amino acid sequence, gene structure and chromosomal location of a novel
 A:Reference number: S32424; MUID:93328631; PMID:8471052
 A:Accession: S32424
 A:Molecule type: DNA
 A:Residues: 1-43, 'T', 45-172, 'N', 174-218 <ZHO>
 A:Cross-references: EMBL:X68676
 A:Note: the authors translated the codon ACG for residue 44 as Ser
 R:Comstock, K.E.; Sanderson, B.J.; Clafilin, G.; Henner, W.D.
 Nucleic Acids Res. 18, 3670, 1990
 A:Title: GST1 gene deletion determined by polymerase chain reaction.
 A:Reference number: 137437; MUID:90301515; PMID:2362832
 A:Accession: 137437
 A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EME
 A:Molecule type: DNA
 A:Residues: 60-118 <RES>
 A:Cross-references: EMBL:X5151; NID:g31922; PIDN:CAA35017.1; PID:g4378955
 R:Singhal, S.S.; Ahmad, H.; Sharma, R.; Gupta, S.; Haque, A.K.; Awasthi, Y.C.
 Arch. Biochem. Biophys. 285, 64-73, 1991
 A:Title: Purification and characterization of human muscle glutathione S-transferases: 6
 A:Reference number: S13905; MUID:91119426; PMID:1846734
 A:Accession: S13905
 A:Molecule type: protein
 A:Residues: 2-15 <SIN>
 A:Experimental source: muscle
 R:Allin, P.; Mannervik, B.; Jornvall, H.
 FEBS Lett. 182, 319-322, 1985
 A:Title: Structural evidence for three different types of glutathione transferase in hum
 A:Reference number: A91336; MUID:8515454; PMID:397955
 A:Accession: B22457
 A:Molecule type: protein
 A:Residues: 2-24 <ALI>
 R:Tsuchida, S.; Makl, T.; Sato, K.
 J. Biol. Chem. 265, 7150-7157, 1990
 A:Title: Purification and characterization of glutathione transferases with an activity
 A:Reference number: A35187; MUID:90237002; PMID:2110160
 A:Accession: D35187
 A:Molecule type: protein
 A:Residues: 2-10, 'X', 12-14, 'X', 16-17, 'X', 19-25 <TSU>
 A:Experimental source: heart
 R:Singhal, S.S.; Saxena, M.; Awasthi, S.; Ahmad, H.; Sharma, R.; Awasthi, Y.C.
 Biochem. Biophys. Acta 1171, 19-26, 1992
 A:Title: Gender related differences in the expression and characteristics of glutathione
 A:Reference number: S27188; MUID:93042004; PMID:1420361
 A:Accession: S27188
 A:Molecule type: protein
 A:Residues: 2-13 <S12>
 A:Experimental source: colon
 R:Pearson, W.R.; Vorachek, W.R.; Xu, S.J.; Berger, R.; Hart, I.; Vannals, D.; Patterson,
 Am. J. Hum. Genet. 53, 220-233, 1993
 A:Title: Identification of class-mu glutathione transferase genes GSTM1-GSTM5 on human C
 A:Reference number: I51867; MUID:93304417; PMID:8317488
 A:Accession: I51867
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 125-172, 'N', 174-186 <RE2>
 A:Cross-references: GB:S62935; NID:9386429; PIDN:AA013938.1; PID:g4261638
 C:Comment: The GSTM1 locus for the mu isoenzyme of glutathione transferase is polymorphi
 stillbene ozide.
 C:Genetics:
 A:Gene: GDB:GSTM1; GSTM16

A:Cross-references: GDB:120020; OMIM:138350
 A:Map position: 1p13.3-1p13.3
 A:introns: 12/3; 38/1; 59/3; 87/1; 120/3; 152/3
 C:Superfamily: glutathione transferase
 C:Keywords: dimer; polymorphism; transferase
 F:2-218/Product: glutathione transferase class mu, GSTM1 #status predicted <Mat>
 Query Match 39.7%; Score 467.5; DB 2; Length 218;
 Best Local Similarity 42.6%; Pred. No. 1.9e-33;
 Matches 92; Conservative 39; Mismatches 80; Indels 5; Gaps 1;
 QY 1 MSPILGMYKIKGLVPTRLLEYLEEYEEHLEERDEG-----DKYRNKFELEFEFNU 55
 DB 1 MPAILGMYKIRGLAHIRLLLETDSEYEKKTTPDADPDYDSQMLNEKFKGLDFPNU 60
 QY 56 PYYIDGVLKLTOSMAIRYIADKNNMLGSCPKERAEISMLEGAVLDIRYGVSRHAYSKDF 115
 DB 61 PYLIDGAHKITQSNAILCYIARKHNLGFEFEKEIKYVDLENGTMDNHQGLMICYNPF 120
 QY 116 ETLKVPFLSKLPMLKMFEDRLCHKTYLNGDHTHDFMLYDALDVLVYLNDFPCLDAFR 175
 DB 121 EKLKPYLEELPEKLYSEFLGKRPWFAGNKITFDVFLYDVLDRIFEPKCLDAFPN 180
 QY 176 IVCFKRIRIPIQIDRYLKSSKRYIAMPLOGQATFG 211
 DB 181 LKDFISRFEGLEKISAYMKSSRFLPRPVFSKMAVWG 216

Search completed: May 20, 2003, 11:39:23
 Job time : 9.41126 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 08:02:28 ; Search time 5.14571 Seconds

(without alignments)
1757.161 Million cell updates/sec

Title: US-10-081-408-4

Perfect score: 1179
Sequence: 1 MSPILGYWKIKGLVQPTRL.....IAPLQGMQATFGGDHPK 218

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1179	100.0	218	1	GR26_SCHUA
2	990	84.0	218	1	GR27_SCHUA
3	927	78.6	218	1	GR26_SCHUA
4	709.5	60.2	220	1	GR29_FASHE
5	699	59.3	217	1	GR27_FASHE
6	681	57.8	217	1	GR28_FASHE
7	680	57.7	217	1	GR26_FASHE
8	481.5	40.8	217	1	GR26_FASHE
9	481.5	40.8	217	1	GR26_FASHE
10	480.5	40.8	218	1	GR26_FASHE
11	478.5	40.6	218	1	GR26_FASHE
12	477.5	40.5	217	1	GR26_FASHE
13	473.5	40.2	217	1	GR26_FASHE
14	473.5	40.2	217	1	GR26_FASHE
15	468.5	38.7	217	1	GR26_FASHE
16	468.5	38.7	217	1	GR26_FASHE
17	467.5	39.7	217	1	GR26_FASHE
18	465.5	39.5	217	1	GR26_FASHE
19	465.5	39.3	217	1	GR26_FASHE
20	461.5	39.1	217	1	GR26_FASHE
21	455.5	38.6	217	1	GR26_FASHE
22	452.5	38.4	217	1	GR26_FASHE
23	448.5	38.0	224	1	GR26_FASHE
24	437.5	37.1	224	1	GR26_FASHE
25	432.5	36.7	219	1	GR26_FASHE
26	385.5	32.7	219	1	GR26_FASHE
27	251.5	21.3	208	1	GR26_FASHE
28	248.5	21.1	208	1	GR26_FASHE
29	246	20.9	210	1	GR26_FASHE
30	244	20.7	209	1	GR26_FASHE
31	242	20.5	207	1	GR26_FASHE
32	240	20.4	209	1	GR26_FASHE
33	238	20.2	209	1	GR26_FASHE

34	237	20.1	209	1	GR26_FASHE	P19157 mus musculus
35	237	20.1	209	1	GR26_FASHE	P04906 rat mus musculus
36	236	20.0	209	1	GR26_FASHE	P46424 cricetus
37	235	19.9	209	1	GR26_FASHE	O28514 macaca mulatta
38	233	19.8	209	1	GR26_FASHE	P46425 mus musculus
39	230	19.5	209	1	GR26_FASHE	P28801 bos taurus
40	219.5	18.6	208	1	GR26_FASHE	P10299 caenorhabditis
41	214.5	18.2	223	1	GR26_FASHE	O08863 oryctolagus
42	204.5	17.3	210	1	GR26_FASHE	P83325 bufo bufo
43	200.5	17.0	221	1	GR26_FASHE	P04903 rat mus musculus
44	198.5	16.8	221	1	GR26_FASHE	P10648 mus musculus
45	198.5	16.8	222	1	GR26_FASHE	P13745 mus musculus

ALIGNMENTS

RESULT 1
GR26_SCHUA STANDARD: PRT: 218 AA.
ID GR26_SCHUA
AC P08515;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Glutathione S-transferase 26 kDa (EC 2.5.1.18) (GST 26) (S26 antigen)
DE (GST class-alpha)
OS Schistosoma japonicum (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoidae;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6182;
[1]
RP SEQUENCE FROM N.A.
RP MEDLINE=87041520; PubMed=3095841;
RA Smith D.B., Davern K.M., Board P.G., Tiu W.U., Garcia E.G.,
RA Mitchell G.F.;
RT "Mr 26,000 antigen is a parasite glutathione S-transferase,"
RT WEHI 129/J mice is a parasite glutathione S-transferase,"
RL Proc. Natl. Acad. Sci. U.S.A. 83:8703-8707(1986).
RN [2]
RP REVIEWS.
RA Smith D.B., Davern K.M., Board P.G., Tiu W.U., Garcia E.G.,
RA Mitchell G.F.;
RL Proc. Natl. Acad. Sci. U.S.A. 84:6541-6541(1987).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=95276631; PubMed=7538846;
RA Lim K., Ho J.X., Keeling K., Gilliland G.L., Ji X., Ruker F.,
RA Carter D.C.;
RT "Three-dimensional structure of Schistosoma japonicum glutathione S-
transferase fused with a six-amino acid conserved neutralizing
epitope of gp41 from HIV.";
RL Protein Sci. 3:2233-2244(1994).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX MEDLINE=95156484; PubMed=7853399;
RA McTigue M.A., Williams D.R., Tainer J.A.;
RT "Crystal structures of a schistosomal drug and vaccine target:
glutathione S-transferase from Schistosoma japonica and its complex
with the leading antischistosomal drug praziquantel.";
RL J. Mol. Biol. 246:21-27(1995).
CC -I- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -I- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
HAEMATIN IN THE PARASITE GUT.
CC -I- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC -I- SUBUNIT: HOMODIMER.
CC -I- MISCELLANEOUS: THERE ARE AT LEAST TWO ISOENZYMES OF GST IN
S. JAPONICUM.
CC -I- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: M14654; AAB59203.1; -
DR PIR: A26484; A26484.
DR PDB: 1GNE: 30-NOV-94.
DR PDB: 1GTA: 07-FEB-95.
DR PDB: 1GTB: 01-DEC-95.
DR PDB: 1B8X: 12-APR-99.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C: 1.
DR Pfam: PF02798; GST_N: 1.
DR Trasnferase; Antigen; Multigene family; 3D-structure.
KW SEQUENCE 218 AA; 25498 MW; 5E2AC18BD0EF13F CRC64;
SQ

Query Match 100.0%; Score 1179; DB 1; Length 218;
Best Local Similarity 100.0%; Pred. No. 1.8e-98;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSPILGYWKIKGLVPTLLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYWKIKGLVPTLLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYYID 60
OY 61 GGVKLTQSMAIIRYADKHNHMGCPKERAETSMLEGAVIDIRGVSRISAKDEETLV 120
DB 61 GGVKLTQSMAIIRYADKHNHMGCPKERAETSMLEGAVIDIRGVSRISAKDEETLV 120
OY 121 DFLSLKPEMLKMFEDRLCHKTYLNGDHYTHPDMFLYDALDVLYNDPMDLAPFLYCFK 180
DB 121 DFLSLKPEMLKMFEDRLCHKTYLNGDHYTHPDMFLYDALDVLYNDPMDLAPFLYCFK 180
OY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOATFGGDDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOATFGGDDHPK 218
OY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOATFGGDDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOATFGGDDHPK 218

RESULT 2

GT27_SCHMA STANDARD; PRT; 218 AA.
ID GT27_SCHMA
AC P35661;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Glutathione S-transferase 26 kDa (EC 2.5.1.18) (GST 26) (SM26/2)
DE antigen) (GST class-alpha).
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6183;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92131046; PubMed-1775156;
RA Wright M.D., Harrison R.A., Melder A.M., Newport G.R., Mitchell G.F.;
RT "Another 26-kilodalton, glutathione S-transferase of Schistosoma
mansoni".

RL Mol. Biochem. Parasitol. 49:177-179(1991).
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
CC HAEMATIN IN THE PARASITE GUT.
CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: M73624; -; NOT_ANNOTATED_CDS.
DR PIR: A45556; A45556.
DR HSRP: P08515; 1GTA.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C: 1.
DR Pfam: PF02798; GST_N: 1.
DR Trasnferase; Antigen; Multigene family.
KW SEQUENCE 218 AA; 25411 MW; DDD3EE9028B36185 CRC64;
SQ

Query Match 84.0%; Score 990; DB 1; Length 218;
Best Local Similarity 82.6%; Pred. No. 1.5e-81;
Matches 180; Conservative 19; Mismatches 19; Indels 0; Gaps 0;

OY 1 MSPILGYWKIKGLVPTLLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MAPKLGWKIKGLVPTLLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYYID 60
OY 61 GGVKLTQSMAIIRYADKHNHMGCPKERAETSMLEGAVIDIRGVSRISAKDEETLV 120
DB 61 GGVKLTQSMAIIRYADKHNHMGCPKERAETSMLEGAVIDIRGVSRISAKDEETLV 120
OY 121 DFLSLKPEMLKMFEDRLCHKTYLNGDHYTHPDMFLYDALDVLYNDPMDLAPFLYCFK 180
DB 121 DFLSLKPEMLKMFEDRLCHKTYLNGDHYTHPDMFLYDALDVLYNDPMDLAPFLYCFK 180
OY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOATFGGDDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOATFGGDDHPK 218
OY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOATFGGDDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMOATFGGDDHPK 218

RESULT 3

GT26_SCHMA STANDARD; PRT; 218 AA.
ID GT26_SCHMA
AC P15964;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Glutathione S-transferase 26 kDa (EC 2.5.1.18) (GST 26) (SM26/1)
DE antigen) (GST class-alpha).
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6183;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-Puerto Rican;
RX MEDLINE-90348716; PubMed-2385266;
RX Troitein F., Kieny M.P., Verwaerde C., Torpier G., Pierce R.J.,
RA Ballou J.-M., Schmitt D., Lecocq J.-P., Capron A.;
RT "Molecular cloning and tissue distribution of a 26-kilodalton
Schistosoma mansoni glutathione S-transferase".
RL Mol. Biochem. Parasitol. 41:35-44(1990).
RN [2]
RP SEQUENCE OF 8-218 FROM N.A.
RX STRAIN-Puerto Rican;
RX MEDLINE-90271935; PubMed-1693415.
RX Henkle K.J., Davern K.M., Wright M.D., Ramos A.J., Mitchell G.F.;
RT "Comparison of the cloned genes of the 26- and 28-kilodalton
glutathione S-transferases of Schistosoma japonicum and Schistosoma
mansoni".

RL Mol. Biochem. Parasitol. 40:23-34(1990).
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF


```

CC      HAEMATIN IN THE PARASITE GUT.
CC      CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC      -1- SUBUNIT: HOMODIMER.
CC      -1- TISSUE SPECIFICITY: TEGUMENT AND IN SUBTEGUMENTARY PARENCHYMAL
CC      CELLS. GST 26 MAY BE ACTIVELY EXCRETED BY ADULT WORMS.
CC      -1- MISCELLANEOUS: THERE ARE AT LEAST THREE ISOENZYMES OF GST IN
CC      S. MANSONI.
CC      -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      DR      EMBL, M31106; AAA29888.1; -.
CC      DR      EMBL, M26913; AAA29889.1; -.
CC      DR      PIR, A45523; A45523.
CC      DR      HSSP; P08515; 1GTA.
CC      DR      InterPro: IPR004046; GST_Cterm.
CC      DR      InterPro: IPR004045; GST_Nterm.
CC      DR      Pfam; PF00043; GST_C; 1.
CC      DR      Pfam; PF02798; GST_N; 1.
CC      KW      Transferase; Antigen; Multigene family.
CC      SQ      SEQUENCE 218 AA; 25401 MW; 061A6548A842D6E8 CRC64;
CC
CC      Query Match 79.6%; Score 927; DB 1; Length 218;
CC      Best Local Similarity 79.8%; Pred. No. 6.3e-76;
CC      Matches 174; Conservative 15; Mismatches 29; Indels 0; Gaps 0.
CC
QY      1 MSPILGKIRKINGVQPTRLLEYLEEKYEHLERDEGDKWKKFELGLEPPNLPYYID 60
Db      1 MAPKRGYKWKVGLVQPTRLLEHLEETYEERAYDRNEIDAMNDKFKLGLLEPPNLPYYID 60
QY      61 GDVKLTGSMATIRYADNHNMGCPKRAETSMLEGAVTDRGVSRAYSKEDETLKV 120
Db      61 GDFKLTGSMATIRYADNHNMGACPKEAEETSMLEGAVTDRMGVLRAYKREYETLV 120
QY      121 DFLSLPEMLKMFEDRLCKTYLNDHYTHPDPMLYDALDVLYMDPMDLAFPKLYCKR 180
Db      121 DFLNKLPERLKMFEDERLSKTYLNGNCVTHPDPMLYDALDVLYMDSCLNPEPKLYSPK 180
QY      181 KRTEAIPQIDKYLKSSKTIAMPLOGSQMOTFGGGRPPK 218
Db      181 KTEEDLPQIKNYLNSRYIKWPLQGWDAITFGGGRPPK 218
CC
RESULT 4
GT29_FASHE
ID      GT29_FASHE STANDARD: PRT; 220 AA.
AC      P36398;
DT      15-DEC-1998 (Rel. 37, Created)
DT      15-DEC-1998 (Rel. 37, Last sequence update)
DT      15-DEC-1998 (Rel. 37, Last annotation update)
DE      Glutathione S-transferase 26 kDa 1 (EC 2.5.1.18) (GST1) (FHL) (GST
DE      class-alpha).
OS      Fasciola hepatica (Liver fluke).
OC      Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
OC      Echinostomidae; Echinostomidae; Fascioloidae; Fasciolidae; Fasciola.
OX      NCBI_TaxID=6192;
RN      [1]
RP      SEQUENCE FROM N.A..
RX      MEDLINE=92155306; Pubmed=1740183;
RA      Panaccio M., Wilson L.R., Crameri S.L., Wjffels G.L., Spithill T.W.;
RT      "Molecular characterization of cDNA sequences encoding glutathione S-
RL      Exp. Parasitol. 74:232-237(1992).
RN      [2]
RP      ERRATUM.
RX      MEDLINE=94039664; Pubmed=8224094;
RA      Panaccio M., Wilson L.R., Crameri S.L., Wjffels G.L., Spithill T.W.;

```

```

RL Exp. Parasitol. 77:385-385(1993).
RN [3]
RP SEQUENCE OF 22-220 FROM N.A.
RA Clement S.:
RC Patent number WO9008819, 09-AUG-1990.
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC PARASITE DETOXIFICATION SYSTEM, OTHER FUNCTIONS ARE ALSO
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
CC HEMATEIN IN THE PARASITE GUT.
CC -1- CATALYTIC ACTIVITY: RX + glutathione = HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; A00993; CA00118.1; -.
DR HSSP; P31670; IHEH.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR004045; GST_Nterm.
DR Pfam; PF00043; GST_C_1.
DR Pfam; PF02798; GST_N_1.
KW Transferrase; Antigen; Multigene family.
FT INITI-MET 0
FT CONFLICT 22
FT CONFLICT 110
FT CONFLICT 189
FT CONFLICT 189
SQ SEQUENCE 220 AA; 25598 MW; 27B9F150B75D101F CRC64;

Query Match 60.2%; Score 709.5; DB 1; Length 220;
Best Local Similarity 60.6%; Pred. No. 1.8e-56;
Matches 129; Conservative 31; Mismatches 52; Indels 1; Gaps 1;

OY 5 LGYWKIKGVQPTRLLELEKEYEENHLEERQEDGKRNKKPELGLPEPNLPYIIDGVK 64
DB 4 LGYWKIRGQGVRLLEL-GEKYEQIYERDGEKWEKRELGDLNPLYYIIDCK 62
OY 65 LTOSAAIIRYADKNHMGCGPKERAISLMGCAVLDIRGVSRAYSKDFETLVKPLS 124
DB 63 LTOSALAIRYADKNGIMGSTPEERARVSMIGCAVLDLRQGLSRISYDPEFOLKEGYIK 122
OY 125 KTEPMKKPEEDLCKHTYNGSDHVPDMELDALDVLVYMDPMCLADAPPKLVCFKKRIE 184
DB 123 DLPTWKMKMSDGLGNPNYIRGTSVSHVDPVYVLEALDAIRLEPRLCHDHPNLDQFRSRIE 182
OY 185 AIPQIDKYLKSSKRYTAMPLOGMATFGGDAPP 217
DB 183 ALPSIKAYMESNRFIKMPNLGNHAGFGGDAPP 215

RESULT 5
GT27_FASHE
AC P31670;
DT 01-JUL-1993 (Rel. 26; Created)
DT 15-DEC-1998 (Rel. 37; Last sequence update)
DT 15-DEC-1998 (Rel. 37; Last annotation update)
DE Glutathione S-transferase 26 kDa 47 (EC 2.5.1.18) (GST47) (FH47)
OS (GST class-alpha).
OS Fasciola hepatica (liver fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
OC Echinostomidae; Echinostomata; Fasciolidae; Fasciola.
OX NCBI-TaxID=6192;
RN [1]
RP SEQUENCE FROM N.A.
MEDLINE=92155306; PubMed=1740183;

```

RA Panaccio M., Wilson L.R., Cramer S.L., Wjffels G.L., Spithill T.W.;
 RT "Molecular characterization of cDNA sequences encoding glutathione S-
 RT transferases of Fasciola hepatica.";
 RN Exp. Parasitol. 74:232-237(1992).
 RN [2]
 RN ERRATUM.
 RX MEDLINE-94039664; PubMed-8224094;
 RA Panaccio M., Wilson L.R., Cramer S.L., Wjffels G.L., Spithill T.W.;
 RN Exp. Parasitol. 77:385-385(1993).
 RN [3]
 RN SEQUENCE OF 7-105 FROM N.A.
 RA Cramer S.;
 RN Patent number WO9008819, 09-AUG-1990.
 RN [4]
 RN X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).
 RX MEDLINE-98035725; PubMed-9367777;
 RA Rossjohn J., Fell S.C., Wilce M.C.J., Sexton J.L., Spithill T.W.,
 RA Parer M.W.;
 RT "Crystallization, structural determination and analysis of a novel
 RT parasite vaccine candidate: Fasciola hepatica glutathione
 RT S-transferase.";
 RN J. Mol. Biol. 273:857-872(1997).
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
 CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
 CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
 CC HAEMATIN IN THE PARASITE GUT.
 CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sib.ch).
 CC -----
 DR EMBL: M7681; AAA29140.1; -;
 DR EMBL: A00996; CAA00121.1; -;
 DR PDB: 1FHE; 29-JUL-98.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR Transferase: Antigen; Multigene family; 3D-structure.
 KW INIT_MER 0
 FT CONFLICT 65 65 T -> I (IN REF. 3).
 FT CONFLICT 102 105 RIGF -> FEEL (IN REF. 3).
 SQ SEQUENCE 217 AA; 25281 MW; 0FB8BBFE63029E03 CRC64;
 Query Match 59.3%; Score 699; DB 1; Length 217;
 Best Local Similarity 58.2%; Pred. No. 1.5e-55;
 Matches 124; Conservative 36; Mismatches 53; Indels 0; Gaps 0;
 QY 5 LGYWKIKGLVOPRLLEYLEEYEEHLERDGDGKRNKKEFLGEPNLPYYIDGDK 64
 DB 4 LGYWKIRGLAOPVRLFLYLEYEEHLVGRDRERKMKSEKFMGDLPLPYIDDKCK 63
 QY 65 LTQSMALIRYIADKHMGLGCPKERAISMLGCAVDIRGVSRIYKNOFELTKYDFLS 124
 DB 64 LTQSVAIMRIADKHMGLGTPPEERARISWIEGAANDLRIGFRVCYNPKFEVKEEYK 123
 QY 125 KLPKMLKMFEDRLCHRTYINGDHVTHPDMLYDALDVLYMDPMCDAPFKLVCFKKRIE 184
 DB 124 ELPKTLKMSDFLGDHRHYLTGSSVSHVDFMLYETLDSIRYLAPHCHDEPFKLEFSRIE 183
 QY 185 AIPQIDKYLKSSKXIAMPLOGWQATFGGCHPP 217
 DB 184 ALPKIKAYMESKRRIKWPINGMAISFGAGDAPP 216

RESULT 6
 ID GT28_FASHE STANDARD: PRT: 217 AA.
 AC P31671;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Glutathione S-transferase 26 kDa 7 (EC 2.5.1.18) (GST) (FW) (GST
 DE class-alpha).
 OS Fasciola hepatica (Liver fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 OC Echinostomida; Echinostomata; Fascioloidae; Fasciola.
 OX NCBI_Taxid-6192;
 RN (1)
 RN SEQUENCE FROM N.A.
 RX MEDLINE-92155306; PubMed-1740183;
 RA Panaccio M., Wilson L.R., Cramer S.L., Wjffels G.L., Spithill T.W.;
 RT "Molecular characterization of cDNA sequences encoding glutathione S-
 RT transferases of Fasciola hepatica.";
 RN Exp. Parasitol. 74:232-237(1992).
 RN [2]
 RN ERRATUM.
 RX MEDLINE-94039664; PubMed-8224094;
 RA Panaccio M., Wilson L.R., Cramer S.L., Wjffels G.L., Spithill T.W.;
 RN Exp. Parasitol. 77:385-385(1993).
 RN [3]
 RN SEQUENCE OF 8-217 FROM N.A.
 RA Cramer S.;
 RN Patent number WO9008819, 09-AUG-1990.
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
 CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
 CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
 CC HAEMATIN IN THE PARASITE GUT.
 CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sib.ch).
 CC -----
 DR EMBL: M7680; AAA29139.1; -;
 DR EMBL: A00994; CAA00119.1; -;
 DR HSP: P31670; 1FHE.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR Transferase: Antigen; Multigene family;
 KW INIT_MER 0
 FT CONFLICT 35 42 NDREKWLG -> MIGNRNGNA (IN REF. 3).
 FT CONFLICT 188 196 IKYEKSER -> SRYMSRA (IN REF. 3).
 SQ SEQUENCE 217 AA; 25196 MW; 00999ELF59E49A49E CRC64;
 Query Match 57.8%; Score 681; DB 1; Length 217;
 Best Local Similarity 58.7%; Pred. No. 6.2e-54;
 Matches 125; Conservative 29; Mismatches 59; Indels 0; Gaps 0;
 QY 5 LGYWKIKGLVOPRLLEYLEEYEEHLERDGDGKRNKKEFLGEPNLPYYIDGDK 64
 DB 4 LGYWKIRGLAOPVRLFLYLEYEEHLVGRDRERKMKSEKFMGDLPLPYIDDKCK 63
 QY 65 LTQSMALIRYIADKHMGLGCPKERAISMLGCAVDIRGVSRIYKNOFELTKYDFLS 124
 DB 64 LTQSVAIMRIADKHMGLGTPPEERARISWIEGAANDLRIGFICYNPKFEELKGDYIK 123

RA Mannervik B., Alin P., Gutenberg C., Jansson H., Tahir M.K.,
 RA Warholm M., Joernvall H.;
 RT "Identification of three classes of cytosolic glutathione transferase
 RT common to several mammalian species: correlation between structural
 RT data and enzymatic properties."; *Proc. Natl. Acad. Sci. U.S.A.* 82:7202-7206(1985).
 RN [7]
 RN CHARACTERIZATION.
 RC STRAIN-CD-1: TISSUE=Liver;
 RX MEDLINE-96189427; PubMed-8605288;
 RA Mitchell A.E., Morin D., Lame M.W., Jones A.D.;
 RT "Purification, mass spectrometric characterization, and covalent
 RT modification of murine glutathione S-transferases."; *J.
 RT Chem. Res. Toxicol.* 8:1054-1062(1995).
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- CATALYTIC ACTIVITY: RX + glutathione = HX + S-glutathione.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- MASS SPECTROMETRY: MW=25838.4; MW_ERR=2; METHOD=Electrospray.
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).
 CC -----
 DR EMBL: J03952; AAA37747.1; -;
 DR EMBL: J04632; AAA37705.1; -;
 DR EMBL: L13448; -; NOT_ANNOTATED_CDS.
 DR EMBL: BC003822; AA03822.1; -;
 DR PIR: A20831; A20831.
 DR PIR: A28946; A28946.
 DR PIR: I24735; I24735.
 DR PIR: A34159; A34159.
 DR PIR: S33860; S33860.
 DR HSSP: P04905; 2CGST.
 DR SMSS-2DPAGE: P10649; MOUSE.
 DR MGI: MGI:95860; Gstm1.
 DR InterPro: IPR004046; GST_cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR InterPro: IPR003081; GST_mu.
 DR Pfam: PF00043; GST_4; 1.
 DR Pfam: PF02798; GST_M; 1.
 DR PRINTS: PRO1267; GSTNSFRASEM.
 DR TRANSFERASE; Multigene family.
 KW INIT_MET 0
 FT SEQUENCE 217 AA; 25839 MW; ALEE3938F590B829 CRC64;
 SQ
 Query Match 40.8%; Score 481.5; DB 1; Length 217;
 Best Local Similarity 44.4%; Pred. NO. 4.3e-36;
 Matches 91; Conservative 38; Mismatches 71; Indels 5; Gaps 1;
 Oy 4 ILGYKINGLVOPTRLLLEVEKEEHLHYENDEG-----DKWRKKKFFLGIEFNNLPY 58
 Db 3 ILGYNVNGLHPIRMALLETYDSDSEKRTYGDAPDFDRSOWLMEKEKLGIDFPLPVL 62
 Oy 59 IDGVKLTQSAIAIRYIAIKHNMGGCKPERAIEISMLGAVLDIYGVSRIRYKDFETL 118
 Db 63 IDGSHKIQNSAIALYLRKHHHLDQETEEERIRADIVENQVMDTMOQLMCLYNDFEQ 122
 Oy 119 KVDPLSKAPEMKPFEDLCHKTYTNGCHVTHPDFMLYDALDVLVYMDPMLDAPFKLVC 178
 Db 123 KPEFKTITPEKKKYSFELGKRPWFAGDKVTVYDFLAVDILDQYRMFEBKCLDAPFNLND 182
 Oy 179 FKKRIEAIPOIDKYLKSSKYTAMPL 203
 Db 183 FLAREGLKISATYMKSSRYTATPI 207

RESULT 9
 GRM1_RAT
 ID GRM1_RAT STANDARD: PRT: 217 AA.
 AC P04905;
 DR 13-AUG-1987 (Rel. 05, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glutathione S-transferase Yb1 (EC 2.5.1.18) (Chain 3) (GST M1-1)
 DE (GST class-mu 1).
 GN GSTM1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (CLONE PGTA200).
 RX MEDLINE-860312882; PubMed-2875437;
 RA Lai H.-C.J., Grove G., Tu C.-P.D.;
 RT "Cloning and sequence analysis of a cDNA for a rat liver glutathione
 RT S-transferase Yb subunit.";
 RL Nucleic Acids Res. 14:6101-6114(1986).
 RN [2]
 RP SEQUENCE FROM N.A. (CLONE PGTA/C44).
 RX MEDLINE-86033768; PubMed-3840477;
 RA Ding G.-F., Lu A.Y.H., Pickett C.B.;
 RT "Rat liver glutathione S-transferases. Nucleotide sequence analysis
 RT of a Yb1 cDNA clone and prediction of the complete amino acid
 RT sequence of the Yb1 subunit.";
 RL J. Biol. Chem. 260:13268-13271(1985).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-86224097; PubMed-3011803;
 RA Ding G.-F., Ding V.D.-H., Rodkey J.A., Bennett C.D., Lu A.Y.H.,
 RA Pickett C.B.;
 RT "Rat liver glutathione S-transferases. DNA sequence analysis of a Yb2
 RT cDNA clone and regulation of the Yb1 and Yb2 mRNAs by
 RT phenobarbital.";
 RL J. Biol. Chem. 261:7952-7957(1986).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-87308179; PubMed-3040722;
 RA Chang C., Saltzman A.G., Sorensen N.S., Hlipakka R.A., Iiao S.;
 RT "Identification of glutathione S-transferase Yb1 mRNA as the
 RT androgen-repressed mRNA by cDNA cloning and sequence analysis.";
 RL J. Biol. Chem. 262:11901-11903(1987).
 RN [5]
 RP SEQUENCE OF 1-23.
 RC STRAIN-Mistar; TISSUE-Olfactory epithelium;
 RX MEDLINE-93277499; PubMed-8503873;
 RA Ben-Arie N., Khen M., Lancet D.;
 RT "Glutathione S-transferases in rat olfactory epithelium:
 RT purification, molecular properties and odorant biotransformation.";
 RL Biochem. J. 292:379-384(1993).
 RN [6]
 RP MUTAGENESIS OF CYS-86.
 RX MEDLINE-91354218; PubMed-1883338;
 RA Hsien J.-C., Huang S.-C., Chen W.-L., Lai Y.-C., Tam M.F.;
 RT "Cysteine-86 is not needed for the enzymic activity of glutathione S-
 RT transferase 3-3.";
 RL Biochem. J. 278:293-297(1991).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RX MEDLINE-93041702; PubMed-1420139;
 RA Ji X., Zhang P., Armstrong R.N., Gilliland G.L.;
 RT "The three-dimensional structure of a glutathione S-transferase from
 RT the mu gene class. Structural analysis of the binary complex of
 RT isoenzyme 3-3 and glutathione at 2.2-A resolution.";
 RL Biochemistry 31:10169-10184(1992).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RA Fu J.-H., Rose J., Tam M.F., Wang B.-C.;
 RT "New crystal forms of a mu-class glutathione S-transferase from rat
 RT liver.";

RL Acta Crystallogr. D 50:219-224(1994).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE-94153886; PubMed-8110735;
 RA Ji X., Johnson W.M., Seeay M.A., Dickert L., Prasad S.M., Ammon H.L.,
 RM Armstrong R.N., Gilliland G.L.;
 RT "Structure and function of the xenobiotic substrate binding site of a
 RT glutathione S-transferase as revealed by x-ray crystallographic
 RT analysis of product complexes with the diastereomers of 9-(S-
 RT glutathionyl)-10-hydroxy-9,10-dihydrophenanthrene.";
 RL Biochemistry 33:1043-1052(1994).
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- FUNCTION: THE OLFACTORY GST MAY BE CRUCIAL FOR THE ACTIVITY OF THE
 CC OLFACTORY PROCESS.
 CC -1- CATALYTIC ACTIVITY: RX + glutathione = HX + R-S-glutathione.
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- MISCELLANEOUS: RAT LIVER GLUTATHIONE S-TRANSFERASES (GST) CONSIST
 CC OF BINARY COMBINATIONS OF THREE MAJOR CLASSES OF SUBUNITS
 CC DESIGNATED: YA (25 KDA), YB (27 KDA) AND YC (28 KDA).
 CC -1- MISCELLANEOUS: YB SUBCLASS SELECTIVELY BINDS STEROID HORMONES.
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).
 CC -----
 DR EMBL: X04229; CAA27811.1; -;
 DR EMBL: M11719; AAA41287.1; -;
 DR EMBL: J02810; AAA41293.1; -;
 DR PIR: A24085; A24085.
 DR PIR: A25510; A25510.
 DR PIR: A29794; A29794.
 DR PIR: S17167; S17167.
 DR PDB: 1GSB; 31-OCT-93.
 DR PDB: 1GSC; 31-OCT-93.
 DR PDB: 2GST; 31-OCT-93.
 DR PDB: 3GST; 31-JAN-94.
 DR PDB: 4GST; 31-OCT-93.
 DR PDB: 5GST; 31-OCT-93.
 DR PDB: 6GST; 08-NOV-96.
 DR PDB: 6GSU; 08-NOV-96.
 DR PDB: 6GSV; 08-NOV-96.
 DR PDB: 6GSW; 08-NOV-96.
 DR PDB: 6GSX; 08-NOV-96.
 DR PDB: 6GSY; 08-NOV-96.
 DR PDB: 5FWG; 27-JAN-99.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR InterPro: IPR003081; GST_mu.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR PRINTS: PRO1267; GSTRNSFRASEM.
 DR TRANSFERASE; Multigene family; 3D-structure; Olfaction.
 FT INIT MET 0 0
 FT MUTAGEN 86 86
 FT CONFLICT 168 168
 FT CONFLICT 198 199
 FT STRAND 2 7
 FT TURN 11 13
 FT HELIX 14 22
 FT TURN 23 24
 FT STRAND 27 32
 FT TURN 37 39
 FT HELIX 43 46
 FT TURN 47 50
 FT STRAND 61 64
 FT TURN 65 66
 C->S: NO CHANGE IN ACTIVITY.
 I -> N (IN REF. 3).
 KS -> NC (IN REF. 2).

FT STRAND 67 70
 FT HELIX 72 82
 FT TURN 83 84
 FT HELIX 90 114
 FT TURN 115 115
 FT TURN 117 118
 FT HELIX 119 128
 FT TURN 129 129
 FT HELIX 130 141
 FT TURN 142 143
 FT TURN 154 154
 FT HELIX 155 169
 FT TURN 171 176
 FT HELIX 178 188
 FT TURN 189 189
 FT HELIX 191 196
 FT TURN 197 198
 FT TURN 200 201
 FT TURN 210 211
 SQ SEQUENCE 217 AA; 25782 MW; 2ACE8D49DA785118 CRC64;
 Query Match 40.8%; Score 481.5; DB 1; Length 217;
 Best local similarity 44.4%; Pred. No. 4,3e-36;
 Matches 91; Conservative 38; Mismatches 71; Indels 5; Gaps 1;
 QY 4 ILGYWKIKGLVPTRLLEYLEEYEDG-----DKRNKKFELGEPNLPY 58
 DB 3 ILGYWVRGITHPIRLLETDSSYEKRYAMDADPYDRSQWLNKFKLGIDFNPYL 62
 QY 59 IDGDVLTOSMAIIRYIADKHNMLGGCPKRAEISMEGAVLIRGVSNRIVASKDETL 118
 DB 63 IDSRSKITOSNATIRYLARKHNLGTEERIRADIYENOVNRQOLIMCTNPDEKQ 122
 QY 119 KYDFLSKLPKMLKFEEDRLCHKTYLNGDHYTHPDMLDALDVLVYMPCLDAFPKLYC 178
 DB 123 KPEFLKTPKMKMLSEFLKRPWFAGDKVTYVDFLAYDLDYHIFEPKCLDAFPKLD 182
 QY 179 FKRRIRAIPOIDKYIKSKRIAMPL 203
 DB 183 FLARFGLKRIISAYMKSRYSRLSTPI 207
 RESULT 10
 GTM4_HUMAN STANDARD; PRT; 218 AA.
 ID GTM4_HUMAN
 AC 003013;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Glutathione S-transferase Mu 4 (EC 2.5.1.18) (GSTM4-4) (GTS-Mu2)
 DE (GST class-mu 4).
 DE GSTM4.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBITaxID=9606;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE-9332467; PubMed-8349586;
 RA Comstock K.E., Johnson K.J., Rifeberry D., Hennen W.D.;
 RT "Isolation and analysis of the gene and cDNA for a human Mu class
 RT glutathione S-transferase, GSTM4.";
 RL J. Biol. Chem. 268:16958-16965(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93228631; PubMed-8471052;
 RA Zhong S., Spurr N.K., Hayes J.D., Wolf C.R.;
 RT "Deduced amino acid sequence, gene structure and chromosomal location
 RT of a novel human class Mu glutathione S-transferase, GSTM4.";
 RL Biochem. J. 291:41-50(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;

QY	56	PYYIDGGVCKLTQSSAIIIRYATDNNHMGCCPKREAAELSMLEGAVALDRIYCVSIIAASKDF	115
Db	61	PYLIDGSHKTYQSSAIIIRYGRKKNLCGFEEREEIRPIDIENRMMDTRIOMGMICYADP	120
QY	116	ETLKVLESLYLPENMLKFEEDRLCHKTYLNDGTHVHPFMYLDALDVLVYNDPMLDAEPK	175
Db	121	EKKRPEELKCLPQOLKRIYSFELGQPPAFADCKITFDALFVLYDVLDQNRMEFPCLDAFPN	180
QY	176	LWCFKKRIEALPOLDKLKSSTKIYAWPLQGMQATFG	211
Db	181	LKQPMARFEGURKISAYMKTSRFLPSPSVYUUKQATWG	216


```

FT INIT_MET 0 0
FT STRAND 2 5
FT STRAND 7 7
FT STRAND 11 12
FT TURN 13 22
FT HELIX 23 24
FT STRAND 27 29
FT STRAND 32 32
FT STRAND 35 35
FT STRAND 38 39
FT STRAND 41 41
FT HELIX 43 49
FT TURN 50 52
FT STRAND 61 64
FT TURN 65 66
FT STRAND 67 70
FT HELIX 72 81
FT TURN 82 84
FT HELIX 90 113
FT TURN 114 115
FT TURN 117 118
FT HELIX 119 127
FT TURN 128 129
FT HELIX 130 141
FT TURN 142 143
FT STRAND 146 146
FT TURN 147 147
FT STRAND 148 149
FT TURN 150 150
FT STRAND 151 151
FT HELIX 154 169
FT HELIX 171 176
FT HELIX 178 189
FT HELIX 191 197
FT TURN 198 198
FT TURN 214 215
SQ SEQUENCE 217 AA; 25613 MW; 25603A909482CA39 CRC64;

```

```

Query Match 40.2%; Score 473.5; DB 1; Length 217;
Best Local Similarity 44.3%; Pred. No. 2.2e-35;
Matches 94; Conservative 35; Mismatches 78; Indels 5; Gaps 1;

```

```

QY 5 LGYKIKGLVOPRLLEYLEEKEEHLIERDEG----DKRRKKFELGLEPFLPYI 59
DB 4 LGYWNIGLASHISILLLEYDSSYEKKYMGADPYDRSQMLNEKFKLGIDFPLPYLI 63
QY 60 DGDVKTOSMAIIRYIADKHNMLGCGCKPERAEISMLGAVIDIRYGSRIYKSDFTLK 119
DB 64 DGTAKITQSNAILRYIARKHNLCSESEKQIREIDILENOFDSRMQAKLCYDDPEFLK 123
QY 120 VDFLSKLPENLKMFEDRLCHRTYLYNGDVTHPDFMLYDALDVLYMPMCIDAPKLVCF 179
DB 124 PEYLOALPEMLKTSQSLGKQPMFLGDKITFVDIADVLERNGVFEPSCLDAFPNLKDF 183
QY 180 KKRLEAIPQIDKYLKSKSYIAMPLOQAOATFG 211
DB 184 ISREGLEKISAYKSSRFLRPVFTKMAVWG 215

```

RESULT 14

```

GTNU_CAVPO STANDARD; PRT; 217 AA.
AC P16413;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1990 (Rel. 15, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE Glutathione S-transferase B (EC 2.5.1.18) (GST B) (GST class-mu).
GN GSTM1.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hysticognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]

```

```

RP SEQUENCE.
RC TISSUE-Liver;
RX MEDLINE-90236961; Pubmed-2332413;
RA Kamei K., Oshino R., Hara S.;
RT "Amino acid sequence of glutathione S-transferase b from guinea pig
  liver.";
RL J. Biochem. 107:111-117(1990).
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
  OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY..
DR HSP: P04905; 2GST..
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C; 1.
DR Pfam: PF02798; GST_N; 1.
DR PRINTS: PR01267; GSTRNSFRASEM.
KW Transferase; Multigene family.
SQ SEQUENCE 217 AA; 25719 MW; D29F7951D4E9365E CRC64;

```

```

Query Match 40.2%; Score 473.5; DB 1; Length 217;
Best Local Similarity 45.1%; Pred. No. 2.2e-35;
Matches 92; Conservative 34; Mismatches 73; Indels 5; Gaps 1;

```

```

QY 5 LGYKIKGLVOPRLLEYLEEKEEHLIERDEG----DKRRKKFELGLEPFLPYI 59
DB 4 LGYWNIGLASHISILLLEYDSSYEKKYMGADPYDRSQMLNEKFKLGIDFPLPYLI 63
QY 60 DGDVKTOSMAIIRYIADKHNMLGCGCKPERAEISMLGAVIDIRYGSRIYKSDFTLK 119
DB 64 DGTAKITQSNAILRYIARKHNLCSESEKQIREIDILENOFDSRMQAKLCYDDPEFLK 123
QY 120 VDFLSKLPENLKMFEDRLCHRTYLYNGDVTHPDFMLYDALDVLYMPMCIDAPKLVCF 179
DB 124 AEFLEGIPIPKKLLFSQFLGKLPWFAKNKLFYVDLADVDOYMLPKCLEAPNLKDF 183
QY 180 KKRLEAIPQIDKYLKSKSYIAMPLO 203
DB 184 ISREGLEKISAYKSSRFLRPKPL 207

```

RESULT 15

```

GTNU_CRIL0 STANDARD; PRT; 217 AA.
AC Q00285;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutathione S-transferase Y1 (EC 2.5.1.18) (Chain 3) (GST class-mu).
OS Cricetus longicaudatus (Long-tailed hamster) (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10030;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91031445; Pubmed-2226437;
RA de Saint Vincent B.R., Hyrien O., Debatisse M., Buttin G.;
RT "Complete cation of mu class glutathione S-transferase genes and an
  adenylate deaminase gene in cotormycin-resistant Chinese hamster
  fibroblasts.";
RL Eur. J. Biochem. 193:19-24(1990).
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
  OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb.sib.ch/announce> or send an email to license@sib.ch).

DR EMBL: X57489; CAA40026.1; -.
DR PIR: S13202; S13202.
DR HSSP: P04905; 2GST.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR InterPro: IPR03081; GST_mu.
DR Pfam: PF00043; GST_C_1.
DR Pfam: PF02798; GST_N_1.
DR PRINTS: PRO1267; GSTRNSEFRASEM.
KW Transferase; Multigene family.
FT INT_MET 0
SQ SEQUENCE 217 AA; 25688 MW; 34D8A8FBDD8627C0 CRC64; BY SIMILARITY

Query Match	39.7%	Score 468.5;	DB 1,	Length 217;
Best Local Similarity	43.4%	Pred. No. 6.3e-35;		
Matches 89;	Conservative 36;	Mismatches 73;	Indels 5;	Gaps 1;

```

OY 4 ILVWKKINGIWOPTBLLLEYLEBEKEEDHUYEDEC-----DKWRKKKFFELGJEFNPLPY 58
Db 3 ILGVNNVNGGLPIPIRLLETYYDSYEKKYITWGADPDSRDSQMLNEKFLGIDFPLPYL 62
OY 59 IDGVVLKNGSMATIRYIADIKHMLGCGPKERAIEISMLEGAVLDIRGVRIAYSKDFETL 118
Db 63 IDGSHKITROSNAILYIRKKNHLCGBETEERIRVYIVQNAQMDITRQGLMCIYNDFEQ 122
OY 119 KVDLSKLPEMLKMFEDRLCHRTYLVNGDHVTHPDMLYDALDVLVLYMDMCLDAPFKLYC 178
Db 123 KPEFLKTIPEKKMKMSEFSGKRPMPAGDVIYLCGFLAYDVLQYOYMFEPKCLDPPNKLMD 182
OY 179 FKRRLEATIQIDKYLKSSKIYAMPL 203
Db 183 FLAREGLKKTISAYMKTSFRLRPI 207

```

Search completed: May 20, 2003, 11:34:59
Job time : 6.14571 secs


```

Query Match          43.5%; Score 512.5; DB 6; Length 218;
Best Local Similarity 47.6%; Pred. No. 2,8e-38;
Matches 99; Conservative 35; Mismatches 69; Indels 5; Gaps 1;

OY 1 MSPILGWYKIKGLVQPTPLLEYLEEKYEHLHYERDEG----DKMRKKKELGLEPNNL 55
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 MPMLIGVYDINGLHAIRALHLEITDNTNEERQYSGVADAPDYDRSQMLNEKFRIGDIPNL 60
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 56 PYYIDGVKLTQSMATIRYIADKHNMLGCGCKERAEISMLEGAVLDIRYGVRIYSKDF 115
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 PYLIDGTGHTKLTQSNAILRYIARKHNKLCGTEEEEMIRVLDILENQVADVRLAMARICYSDF 120
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 116 ETLVWDFLSKIPENMLKMEEDRLCKHTYTLNGDHYVHPDMLVDALDLYVMQMDLAPPK 175
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 EKLKRGFLKEIKPEKTKSEFLGKRPWAGDKLTVDFLVYDVLDMHRIEFPKCDAPFN 180
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 176 LVCFKKRIEAPQIDDKYLSSKRYIAMPL 203
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 LKDFISREFGLKRIISAYMKSSRFLRGPL 208
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
O90WM9 PRELIMINARY; PRT; 219 AA.
AC O90WM9:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last, sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Glutathione S-transferase (EC 2.5.1.18).
GN GSTM1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_Taxid=8355;
[1]
RP SEQUENCE FROM N.A.
RA De Luca A., Favaloro B., Sacchetta P., Angelucci S., Di Ilio C.;
RT "Molecular cloning, expression and site-directed mutagenesis of a
RL liver Mu-class glutathione S-transferase from Xenopus laevis.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; A0416998; CADD1094.1; -
DR InterPro; IPR004046; GST_Cterm.
DR Pfam; PF00043; GST_C; 1.
DR Pfam; PF02798; GST_N; 1.
KW Transferase.
FT CHAIN 2 219 XLGSTM1.
SQ SEQUENCE 219 AA: 25359 MW: 567F8AEDDA82DFD CRC64;
```

```

Query Match 505.5; DB 13; Length 219;
Best Local Similarity 47.2%; Pred No. 1,2e-37;
Matches 102; Conservative 33; Mismatches 76; Indels 5; Gaps 1
42.9%; Score 505.5; DB 13; Length 219;
Pred No. 1,2e-37;
Matches 102; Conservative 33; Mismatches 76; Indels 5; Gaps 1

```

```

ID      097117;      PRELIMINARY;      PRT:      223 AA.
AC      097117;
DT      01-MAY-1999 (TREMBLrel. 10, Created)
DT      01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE      Glutathione S-transferase.
OS      Boophilus microplus (Cattle tick).
OC      Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC      Parasitiformes; Ixodida; Ixodidae; Boophilus.
OX      NCBI_TaxID=6941;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=99381232; PubMed=10451925;
RA      He H., Chen A.-C., Davey R.B., Iyie G.W., George J.E.;
RT      "Characterization and molecular cloning of a glutathione S-transferase
RL      gene from the tick, Boophilus microplus (Acari: Ixodidae).";
RL      Insect Biochem. Mol. Biol. 29:737-743(1999).
DR      EMBL; AF077609; AADI5991.1; -.
DR      HSSP; P20136; 1GSU.
DR      InterPro; IPR004046; GST_Cterm.
DR      InterPro; IPR003081; GST_mu.
DR      InterPro; IPR004045; GST_Nterm.
DR      Pfam; PF00043; GST_C; 1.
DR      Pfam; PF02798; GST_N; 1.
DR      PRINTS; PR01267; GSTRNSPBASEM.
KW      Transferase
SQ      SEQUENCE 223 AA; 25604 MW; F276ACE2D8259174 CRC64;

```

```

Query Match          42.6%; Score 502.5; DB 5; Length 223;
Best Local Similarity 47.2%; Pred.No.2,3e-37;
Matches 102; Conservative 29; Mismatches 80; Indels 5; Gaps 1.

OY      1 MSPILGYMKIKCGLVPTKLLLEYLEEKYEAEHLERYD-----EGDKMRNKKFELGLEFPML 55
         |||::|||::||| ||| ||| :: : | : | : ||| ||| |||
DB       1 NAPIVGWYDIRLADPPIRILLAAHVAKAYDDKRYTCGPPDPFRSSWLNETKTLGLEFPNL 60
OY      56 PVIIDGVKLTQSMATIRYIADKNHNLGGCPKERAEISMLEGAVLDIRGVSRIVASYKDF 115
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB       61 PVIIDGVKLTQSMATIRLARKHGLEGTEAEKQKVVDSEGOAFDRNMWRICNPVF 120
OY      116 ETLKVDFTLSKLPEMLKMEEDRLCHKTYYLNGDHVTHPDFMLYDALDVLYLMDPYCLDAFK 175
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB       121 EKTKGDYELNPLPASLKAFESDYLGTHKFEGDNITLYDEFAIYEMLAQHILFAPDCLEDFAN 180
OY      176 LVCFFKRTEALPQIDKYLLSKSYKIAMPLOGMATEG 211
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB       181 LKAFLVRIEALPHVAAYLKSCKTKPWLNGBMASFG 216

RESULT 7
OBR516 ID OBR516 PRELIMINARY: PRT: 218 AA.
AC OBR516:
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE Glutathione transferase GSTW7-7 (CC 2.5.1.18).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090; [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Jr.
RA Guo J., Zimniak L., Zimniak P., Orchard J.L., Singh S.V.:
RT "Cloning and expression of a novel mu class murine glutathione
RT transferase isoenzyme."
RL Submitted (JAN-2002) to the EMBL/GenBank/DDJB databases.
DR EMBL; AF464943; AAL76248.1; -.
KW Transferrase.
SQ SEQUENCE 218 AA; 25519 MW; 81F15DBB46118102 CRC64;
```


RT 11ver.":
 RL Toxicol. Sci. 56:26-36(2000).
 DR EMBL: AF2007710; AAF08540.1; -.
 DR HSSP: P28161; 1HNA.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR003081; GST_mu.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR PRINTS: PRO1267; GSTRNSFRASEM.
 DR Transferrase.
 KW SEQUENCE 218 AA; 25708 MW; F2E509C3949F9051 CRC64;

Query Match 40.2%; Score 474.5; DB 5; Length 218;
 Best Local Similarity 42.4%; Pred. No. 7.3e-35;
 Matches 96; Conservative 34; Mismatches 81; Indels 5; Gaps 1;

OY 1 MSPILGWKIKGLVOPTRLLLEYLEKEYEHLERDEG-----DKWRNKKFELGLEFPNL 55
 DB 1 MPMTGLWNRIRGLASIRLLLEYTGSSYEKKYTMGDADYDRSOWLNKFKLGIDFPNL 60
 OY 56 PYIDGDVKLTQSMATIRYIADKHNMLGCCPKERAISMLEGAVLDIRYGSRIAYSKDF 115
 DB 61 PYLIDGTHKITQSNALIRYIARKHNLCGETEKEKIRELIDENQIMDMRMLARLCYDPDF 120
 OY 116 ETLKVDLSKLPENLMKMFEDRLCHKTYLNGDHTHPDMLYDALDVLVLMDCPLDAFPK 175
 DB 121 EKLKPEYLEGLPEMLKLYSQFLGDKITFVDFIAYDLERNQVFEPSCLDAFPN 180
 OY 176 LVCFKKRIEAIPOIDKYLKSKYIAMPLOGMOATFG 211
 DB 181 LKDFISREBGLERISAYMKSSRLPRPVFTKMAVWG 216

RESULT 11
 ID 027653 PRELIMINARY; PRT; 219 AA.
 AC 027653;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Glutathione transferase (EC 2.5.1.18).
 GN EMGST1.
 OS Echinococcus multilocularis.
 OC Eukaryota; Metazoa; Platyhelminthes; Cestoda; Eucestoda;
 OC Cyclophyllidae; Taeniidae; Echinococcus.
 OX NCBI_Taxid=6211;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96379220; PubMed=8784771;
 RA Liebau E., Muller V., Lucius R., Walter R.D., Henkle-Duhrsen K.;
 RT "Molecular cloning, expression and characterization of a recombinant
 glutathione S-transferase from Echinococcus multilocularis";
 RL Mol. Biochem. Parasitol. 77:49-56(1996).
 DR EMBL: X85736; CAA59739.1; -.
 DR HSSP: P20136; 1GST.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR Transferrase.
 KW SEQUENCE 219 AA; 25470 MW; 74FA820AC56F745B CRC64;

Query Match 40.2%; Score 474.5; DB 5; Length 219;
 Best Local Similarity 42.4%; Pred. No. 7.4e-35;
 Matches 92; Conservative 40; Mismatches 80; Indels 5; Gaps 1;

OY 1 MSPILGWKIKGLVOPTRLLLEYLEKEYEHLVE-----RDEGKWNKKFELGLEFPNL 55
 DB 1 MAPPLAWMDIRGLAEOSRLKLTLEYEYDDKRYKIGSAPFDRSAMLSEKFSGLDFPNL 60
 OY 56 PYIDGDVKLTQSMATIRYIADKHNMLGCCPKERAISMLEGAVLDIRYGSRIAYSKDF 115
 DB 61 PYLIDGTHKITQSNALIRYIARKHNLCGETEKEKIRELIDENQIMDMRMLARLCYDPDF 120

DB 61 PYIDGDVKLTQSMATIRYIADKHNMLGCCPKERRARVLMHLOCEVVDLRMAFTRTCYSPDF 120
 OY 116 ETLKVDLSKLPENLMKMFEDRLCHKTYLNGDHTHPDMLYDALDVLVLMDCPLDAFPK 175
 DB 121 EKLKPEYLEGLPEMLKLYSQFLGDKITFVDFIAYDLERNQVFEPSCLDAFPN 180
 OY 176 LVCFKKRIEAIPOIDKYLKSKYIAMPLOGMOATFG 212
 DB 181 LKAYLSREBGLERISAYMKSSRLPRPVFTKMAVWG 217

RESULT 12
 ID 016058 PRELIMINARY; PRT; 219 AA.
 AC 016058;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Glutathione S-transferase.
 GN GSTMU.
 OS Echinococcus granulosus.
 OC Eukaryota; Metazoa; Platyhelminthes; Cestoda; Eucestoda;
 OC Cyclophyllidae; Taeniidae; Echinococcus.
 OX NCBI_Taxid=6210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Fernandez V., Zaha A., Fernandez C.;
 RL Submitted (May-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Fernandez Mancebo V., Chalar C., Martinez C., Zaha A.,
 RA Fernandez Granja C.;
 RT "EGST: a glutathione S-transferase gene from Echinococcus
 granulosus";
 RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF005928; AAB6318.1; -.
 DR EMBL: AF101269; AAD16438.1; -.
 DR HSSP: P20136; 1GST.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR Transferrase.
 KW SEQUENCE 219 AA; 25553 MW; 8883E70AD075D154 CRC64;

Query Match 40.2%; Score 473.5; DB 5; Length 219;
 Best Local Similarity 42.4%; Pred. No. 9e-35;
 Matches 92; Conservative 40; Mismatches 80; Indels 5; Gaps 1;

OY 1 MSPILGWKIKGLVOPTRLLLEYLEKEYEHLVE-----RDEGKWNKKFELGLEFPNL 55
 DB 1 MAPPLAWMDIRGLAEOSRLKLTLEYEYDDKRYKIGSPTFDRSAMLSEKFSGLDFPNL 60
 OY 56 PYIDGDVKLTQSMATIRYIADKHNMLGCCPKERAISMLEGAVLDIRYGSRIAYSKDF 115
 DB 61 PYIDGDVKLTQSMATIRYIADKHNMLGCCPKERRARVLMHLOCEVVDLRMAFTRTCYSPDF 120
 OY 116 ETLKVDLSKLPENLMKMFEDRLCHKTYLNGDHTHPDMLYDALDVLVLMDCPLDAFPK 175
 DB 121 EKLKPEYLEGLPEMLKLYSQFLGDKITFVDFIAYDLERNQVFEPSCLDAFPN 180
 OY 176 LVCFKKRIEAIPOIDKYLKSKYIAMPLOGMOATFG 212
 DB 181 LKAYLSREBGLERISAYMKSSRLPRPVFTKMAVWG 217

RESULT 13
 ID 091Y83 PRELIMINARY; PRT; 218 AA.
 AC 091Y83;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)


```

DE Glutathione S-transferase subunit gyc.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriocognathi; Cavidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HARTLEY; TISSUE=LIVER;
RX MEDLINE=98297271; PubMed=9633615;
RA Hiratsuka A., Ogura K., Fujioke H., Sakamoto Y., Okuda H., Wada K.,
RA Tanaka T., Nishiyama T., Matabe T.;
RT "Guinea pig liver Mu-class glutathione S-transferase M1-2 cross-reacts
RT with antibodies to both rat Mu- and theta-class glutathione S-
RT transferases."
RL Arch. Biochem. Biophys. 354:188-196(1998).
DR EMBL: AB000488; BAB47185.1;
DR InterPro: IPR004046; GST_Cterm.
DR Pfam: PF00043; GST_C; 1.
DR Pfam: PF02798; GST_N; 1.
DR Pfam: PF02798; GST_N; 1.
KM Transferase.
SQ SEQUENCE 218 AA; 25695 MW; D2B9E31F9FEF1B8 CRC64;

Query Match 40.0%; Score 471.5; DB 11; Length 218;
Best Local Similarity 44.7%; Pred. No. 1.4e-34;
Matches 93; Conservative 33; Mismatches 77; Indels 5; Gaps 1;

OY 1 MSPIIGYMKIKGLVOTRLLLEYLEKYEHLHYERDEG-----DKRRNKFELGLEFPNL 55
DB 1 MPTTJGVMWIRGLTHAIRLLLEYTDTSYEKRYTWGDADFTSQWLNRKFKGLDFPNL 60
OY 56 PYIDGDVVLQTOSMAIRYIADKHNMLGCGPKERAETSMLEGAVLDIRYGSRIAYSKDF 115
DB 61 PVLIDGTHLQTSNMLRIARLNKNCSTEEERIMDLLENQVDFRQLQGNVCTSPDF 120
OY 116 ETLKVDFLSKLPMLKMFEDRLCHKTYLNGDHYHDFMLYDALDVLVYMDPMDCLDAFPR 175
DB 121 EKKKEPEFLKSIPEKMKIYSEFLGKRPWAGDKVTYVDYFLAYDLDDYRMFEPECLDAFPR 180
OY 176 LVCFKKRIAPQIDKYLKSSKTYIAMPL 203
DB 181 LKDFLARFELGLKRIKISAYMKSSSFLPRPV 208

RESULT 14
O9WU21 PRELIMINARY; PRT; 218 AA.
AC O9WU21.
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Glutathione S-transferase Yb4 (EC 2.5.1.18).
GN GSTRYB4.
OS Rattus norvegicus (rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=88298790; PubMed=3403534;
RA Lai H.C., Qian B., Grove G., Tu C.P.;
RT "Gene expression in the evolution of the yb multigene family."
RT "Conversion in the evolution of the yb multigene family."
RL J. Biol. Chem. 263:11389-11395(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RA Qian B., Tu C.P.;
RT "The Rat Glutathione S-transferase Yb4 genomic clone sequence."
RL EMBL: AF106861; AAD22630.1;
DR EMBL: AF106861; AAD22630.1;
DR HSSP: P04905; 2GST.

```

```

DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C; 1.
DR Pfam: PF02798; GST_N; 1.
DR Pfam: PF02798; GST_N; 1.
KM Transferase.
SQ SEQUENCE 218 AA; 25643 MW; 4E9C22P9AEAC6C4 CRC64;

Query Match 39.4%; Score 464.5; DB 11; Length 218;
Best Local Similarity 43.3%; Pred. No. 5.8e-34;
Matches 90; Conservative 35; Mismatches 78; Indels 5; Gaps 1;

OY 1 MSPIIGYMKIKGLVOTRLLLEYLEKYEHLHYERDEG-----DKRRNKFELGLEFPNL 55
DB 1 MPTTJGVMWIRGLTHAIRLLLEYTDTSYEKRYTWGDADFTSQWLNRKFKGLDFPNL 60
OY 56 PYIDGDVVLQTOSMAIRYIADKHNMLGCGPKERAETSMLEGAVLDIRYGSRIAYSKDF 115
DB 61 PVLIDGTHLQTSNMLRIARLNKNCSTEEERIMDLLENQVDFRQLQGNVCTSPDF 120
OY 116 ETLKVDFLSKLPMLKMFEDRLCHKTYLNGDHYHDFMLYDALDVLVYMDPMDCLDAFPR 175
DB 121 EKKKEPEFLKSIPEKMKIYSEFLGKRPWAGDKVTYVDYFLAYDLDDYRMFEPECLDAFPR 180
OY 176 LVCFKKRIAPQIDKYLKSSKTYIAMPL 203
DB 181 LKDFLARFELGLKRIKISAYMKSSSFLPRPV 208

RESULT 15
O9DD25 PRELIMINARY; PRT; 218 AA.
AC O9DD25.
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 061005A07Rik protein.
GN RPRGRI1 OR 061005A07Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo N.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Wittlake C., Wilming L.,
RA Wyshnaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL: AK002213; BAB21939.1;
DR HSSP: P28161; 2GTU
DR MGD: MGI:1915562; 0610005A07Rik.
DR MGD: MGI:1932134; Rprgri1.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR003081; GST_mu.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C; 1.

```


GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 07:58:08 ; Search time 18.4058 Seconds

(without alignments)
1578.232 Million cell updates/sec

Title: US-10-081-408-5

Perfect score: 1164

Sequence: 1 MSPILGWTIKIGLVQPTRL.....IAMPDQGQATFGGDHPK 218

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /_Geneseq 101002:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1149	98.7	218	20	AAW84205
2	1149	98.7	218	23	ABBO9809
3	1149	98.7	224	17	AAW02122
4	1149	98.7	225	17	AAW02123
5	1149	98.7	232	16	AAW72793
6	1149	98.7	240	20	AAW3909
7	1149	98.7	245	22	AAW31612
8	1149	98.7	247	20	AAW74114
9	1149	98.7	247	20	AAW74120
10	1149	98.7	247	20	AAW74103

11	1149	98.7	248	20	AAW74112
12	1149	98.7	248	20	AAW74113
13	1149	98.7	248	20	AAW74121
14	1149	98.7	248	20	AAW74122
15	1149	98.7	248	20	AAW74100
16	1149	98.7	248	20	AAW74101
17	1149	98.7	248	20	AAW74102
18	1149	98.7	249	20	AAW74115
19	1149	98.7	252	13	AAW22139
20	1149	98.7	252	13	AAW22460
21	1149	98.7	252	13	AAW22140
22	1149	98.7	257	20	AAW74111
23	1149	98.7	257	20	AAW74119
24	1149	98.7	257	20	AAW74124
25	1149	98.7	257	20	AAW74126
26	1149	98.7	257	20	AAW74105
27	1149	98.7	257	20	AAW74109
28	1149	98.7	258	20	AAW74117
29	1149	98.7	258	20	AAW74104
30	1149	98.7	259	20	AAW74130
31	1149	98.7	259	20	AAW74118
32	1149	98.7	259	20	AAW74125
33	1149	98.7	259	20	AAW74110
34	1149	98.7	262	20	AAW74128
35	1149	98.7	264	20	AAW74129
36	1149	98.7	267	20	AAW74106
37	1149	98.7	271	13	AAW21703
38	1149	98.7	272	19	AAW56327
39	1149	98.7	277	20	AAW74116
40	1149	98.7	277	20	AAW74098
41	1149	98.7	277	20	AAW74099
42	1149	98.7	277	20	AAW74107
43	1149	98.7	279	20	AAW74108
44	1149	98.7	279	20	AAW74097
45	1149	98.7	282	19	AAW56326

ALIGNMENTS

AAW84205	standard: Protein; 218 AA.
AAW84205;	
25-MAR-1999	(first entry)
Protein comprising a purification tag.	
Proteinase site; bone morphogenetic fusion protein; bone binding site;	
bone morphogenetic protein; transforming growth factor beta;	
active fragment; wound healing; bone growth; purification tag.	
Unidentified.	
10-DEC-1998.	
WO9855137-A1.	
03-JUN-1997;	97US-0868452.
02-JUN-1998;	98WO-US11189.
03-JUN-1997;	97US-0868452.
(HALL/) HALL F L.	
(HAMB/) HAN B.	
(NIMN/) NIMNI M E.	
(SHOR/) SHORS E C.	
(WDL/) WU L.	
Hall FL, Han B, Nimi ME, Shors EC, Wu L;	
WPI: 1999-059875/05.	


```

XX Example 11; Page 31-32; 47pp; English.
PS
XX A glutathione-s-transferase epitope (AAW02122) was used to construct
CC fusion proteins with human SH2 domains from fyn (see also AAW02124),
CC Grb2 (AAW02127) and SH-PTP2 (AAW02125); a similar epitope (AAW02123) was
CC used to construct a fusion protein with human p85 SH2 domain (see
CC also AAW02126). These fusion proteins, and tagged proteins including
CC other SH2 domains (see also AAW02119-21), can be used in binding
CC assays to determine the specificity of cpds. To inhibit SH2 domains;
CC cpds. that selectively inhibit the human src SH2 domain are useful
CC in treating bone resorption diseases such as osteoporosis.
XX
SQ Sequence 224 AA;
Query Match 98.7%; Score 1149; DB 17; Length 224;
Best Local Similarity 98.6%; Pred. No. 6.9e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 MSPILGYMKIKGLVOPTRLLLEYLEEKYEENLYERDEGDKMKNKKFELGLEFPNLPYYTD 60
DB 1 MSPILGYMKIKGLVOPTRLLLEYLEEKYEENLYERDEGDKMKNKKFELGLEFPNLPYYTD 60
OY 61 GGVKLTQSMATIRYIADKHNMLGSGPKERAETISMEGAVLDIRYGVSRAYSKEDETLKV 120
DB 61 GGVKLTQSMATIRYIADKHNMLGSGPKERAETISMEGAVLDIRYGVSRAYSKEDETLKV 120
OY 121 DFLSKLPENLKMFEEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPMCLDAFPKLVCFK 180
DB 121 DFLSKLPENLKMFEEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPMCLDAFPKLVCFK 180
OY 121 DFLSKLPENLKMFEEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPMCLDAFPKLVCFK 180
DB 121 DFLSKLPENLKMFEEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPMCLDAFPKLVCFK 180
OY 181 KRIEAIPOIDKYLKSSKTYAMPLOGWQATFGGSDHPK 218
DB 181 KRIEAIPOIDKYLKSSKTYAMPLOGWQATFGGSDHPK 218

RESULT 4
AAW02123
ID AAW02123 standard; Protein: 225 AA.
XX
AC AAW02123;
XX
DT 28-OCT-1996 (first entry)
XX
DE Glutathione-s-transferase epitope.
XX
KW Bone resorption disease; osteoporosis; src SH2 domain antagonist;
KW src homology 2 domain; glutathione-s-transferase; p85 SH2.
XX
OS Not specified.
XX
PN EP727211-A1.
XX
PD 21-AUG-1996.
XX
PF 07-FEB-1996; 96EP-0200270.
XX
PR 29-DEC-1995; 95US-0580868.
PR 10-FEB-1995; 95US-0386381.
PR 07-MAR-1995; 95US-0400220.
PR 30-JUN-1995; 95US-0497357.
PR 11-OCT-1995; 95US-0541080.
XX
PA (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI Dunnington DJ;
XX
DR WPI: 1996-372674/38.
XX
PT Use of selective src SH2 domain ligand - to prepare medicament for
XX treating bone resorption disease
XX
PS Example 11; Page 31-32; 47pp; English.

```

```

XX
CC A glutathione-s-transferase epitope (AAW02122) was used to construct
CC fusion proteins with human SH2 domains from fyn (see also AAW02124),
CC Grb2 (AAW02127) and SH-PTP2 (AAW02125); a similar epitope (AAW02123) was
CC used to construct a fusion protein with human p85 SH2 domain (see
CC also AAW02126). These fusion proteins, and tagged proteins including
CC other SH2 domains (see also AAW02119-21), can be used in binding
CC assays to determine the specificity of cpds. To inhibit SH2 domains;
CC cpds. that selectively inhibit the human src SH2 domain are useful
CC in treating bone resorption diseases such as osteoporosis.
XX
SQ Sequence 225 AA;
Query Match 98.7%; Score 1149; DB 17; Length 225;
Best Local Similarity 98.6%; Pred. No. 6.9e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 MSPILGYMKIKGLVOPTRLLLEYLEEKYEENLYERDEGDKMKNKKFELGLEFPNLPYYTD 60
DB 1 MSPILGYMKIKGLVOPTRLLLEYLEEKYEENLYERDEGDKMKNKKFELGLEFPNLPYYTD 60
OY 61 GGVKLTQSMATIRYIADKHNMLGSGPKERAETISMEGAVLDIRYGVSRAYSKEDETLKV 120
DB 61 GGVKLTQSMATIRYIADKHNMLGSGPKERAETISMEGAVLDIRYGVSRAYSKEDETLKV 120
OY 121 DFLSKLPENLKMFEEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPMCLDAFPKLVCFK 180
DB 121 DFLSKLPENLKMFEEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPMCLDAFPKLVCFK 180
OY 121 DFLSKLPENLKMFEEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPMCLDAFPKLVCFK 180
DB 121 DFLSKLPENLKMFEEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPMCLDAFPKLVCFK 180
OY 181 KRIEAIPOIDKYLKSSKTYAMPLOGWQATFGGSDHPK 218
DB 181 KRIEAIPOIDKYLKSSKTYAMPLOGWQATFGGSDHPK 218

RESULT 5
AAR72793
ID AAR72793 standard; Protein: 232 AA.
XX
AC AAR72793;
XX
DT 16-OCT-1995 (first entry)
XX
DE Glutathione-s-transferase.
XX
KW SDR-1; senescent cell-derived inhibitor; DNA synthesis;
KW senescence; cell proliferation; cancer; therapeutic; vulnery;
KW fusion protein; glutathione-s-transferase.
XX
OS Schistosoma japonicum.
XX
PN WO9506415-A.
XX
PD 09-MAR-1995.
XX
PF 26-AUG-1994; 94WO-US09700.
XX
PR 13-JUL-1994; 94US-0274535.
PR 30-AUG-1993; 93US-0113372.
PR 17-NOV-1993; 93US-0153564.
PR 03-JAN-1994; 94US-0160814.
PR 25-FEB-1994; 94US-0203535.
PR 15-APR-1994; 94US-0228420.
PR 30-JUN-1994; 94US-0268439.
XX
PA (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI Smith JR;
XX
DR WPI: 1995-131101/17.
DR N-PSDB; AAO86774.
XX
PT Nucleic acid encoding a protein or polypeptide that inhibits DNA
XX synthesis in a recipient cell - useful to inhibit cell

```

PT proliferation in tumour cells, treat wound or burn tissue, or as
 PT an antiviral or antiparasitic agent
 XX
 PS Disclosure; Page 130; 169pp; English.
 XX
 CC The senescent cell-derived inhibitor-1 (SDI-1) inhibits DNA
 CC synthesis in a recipient cell and is capable of associating with a
 CC cyclin (esp. cyclin D1) or cyclin-dependent kinase (esp. CDK2).
 CC Production of a fusion protein of SDI-1 and glutathione-S-transferase
 CC (AAR77793) facilitated the preparation and isolation of recombinant
 CC SDI-1.
 CC
 XX
 SQ Sequence 232 AA;
 Query Match 98.7%; Score 1149; DB 16; Length 232;
 Best Local Similarity 98.6%; Pred. No. 7.2e-113;
 Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSPILGKIRKGLVOPTRLLLELEKEEHLERDECDKRNKKFELGLEPNNLPYYID 60
 DB 1 MSPILGKIRKGLVOPTRLLLELEKEEHLERDECDKRNKKFELGLEPNNLPYYID 60
 QY 61 GDVKLTQSMALIRYADKHNMLGGSPKERAETSMLEGAVIDIRYGSRIASKDEETLKV 120
 DB 61 GDVKLTQSMALIRYADKHNMLGGSPKERAETSMLEGAVIDIRYGSRIASKDEETLKV 120
 QY 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPCLDAFPKLVGFK 180
 DB 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLYMDPCLDAFPKLVGFK 180
 QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMQATFGGSDHPK 218
 DB 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMQATFGGSDHPK 218
 RESULT 6
 AAW73909
 ID AAW73909 standard; Protein: 240 AA.
 XX
 AC AAW73909;
 XX
 DT 13-APR-1999 (first entry)
 XX
 DE GST protein sequence.
 XX
 KM Cell surface nonexpressive functional protein; cell-permeating protein;
 KM protein screening; protein production; GST; glutathione-S-transferase.
 XX
 OS Schistosoma japonicum.
 XX
 PN JP11009278-A.
 XX
 PD 19-JAN-1999.
 XX
 PF 23-JUN-1997; 97JP-0165788.
 XX
 PR 23-JUN-1997; 97JP-0165788.
 XX
 PA (SUME) SUMITOMO ELECTRIC IND CO.
 XX
 DR WPI: 1999-145895/13.
 DR N-PSDB: AAX01312.
 XX
 PT Microbe having a nonexpressive functional protein on its surface -
 PT for screening and large-scale preparation of a selected function
 PT protein
 XX
 PS Disclosure; Page 8-9; 13pp; Japanese.
 CC This sequence is the glutathione-S-transferase (GST) protein, which
 CC was expressed in the microbe of the invention. The microbe has a cell
 CC surface nonexpressive functional protein on its surface, fused to a
 CC support protein derived from a cell-permeating protein. The microbe can

CC be used for screening for biological activity of a cell surface
 CC nonexpressive protein. It can also be used for preparation of a selected
 CC functional protein in a large amounts.
 XX
 SQ Sequence 240 AA;
 Query Match 98.7%; Score 1149; DB 20; Length 240;
 Best Local Similarity 98.6%; Pred. No. 7.6e-113;
 Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSPILGKIRKGLVOPTRLLLELEKEEHLERDECDKRNKKFELGLEPNNLPYYID 60
 DB 1 MSPILGKIRKGLVOPTRLLLELEKEEHLERDECDKRNKKFELGLEPNNLPYYID 60
 QY 61 GDVKLTQSMALIRYADKHNMLGGSPKERAETSMLEGAVIDIRYGSRIASKDEETLKV 120
 DB 61 GDVKLTQSMALIRYADKHNMLGGSPKERAETSMLEGAVIDIRYGSRIASKDEETLKV 120
 QY 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPCLDAFPKLVGFK 180
 DB 121 DFLSKLPEMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLYMDPCLDAFPKLVGFK 180
 QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMQATFGGSDHPK 218
 DB 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMQATFGGSDHPK 218
 RESULT 7
 AAB31612
 ID AAB31612 standard; Protein: 245 AA.
 XX
 AC AAB31612;
 XX
 DT 30-APR-2001 (first entry)
 XX
 DE Amino acid sequence of glutathion-S-transferase (GST).
 XX
 KM Heat shock protein; Hsp: Th1 response; Th1 cell; CD4+ T lymphocyte cell;
 KM lymphocyte; Hsp65; Hsp40; Hsp10; Hsp60; Hsp71; microbial pathogen;
 KM glutathion-S-transferase; GST.
 XX
 OS unidentified.
 XX
 PN WO200104344-A2.
 XX
 PD 18-JAN-2001.
 XX
 PF 10-JUL-2000; 2000WO-US18828.
 XX
 PR 08-JUL-1999; 99US-0143757.
 XX
 PA (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
 XX
 PI Siegel M, Chu NR, Mizzzen LA;
 XX
 DR WPI: 2001-138361/14.
 DR N-PSDB: AAF25015.
 XX
 PT Screening for compounds that stimulate Th1-like responses in CD4+ T
 PT lymphocyte cells -
 XX
 PS Example 9; Fig 8; 88pp; English.
 CC The present sequence represents a glutathion-S-transferase (GST), which
 CC was used linked to HPV16 E7 protein, for purification purposes. The E7
 CC protein was used to construct a fusion protein with Mycobacterium bovis
 CC heat shock proteins (Hsp). The fusion proteins are used in the method of
 CC the invention. The specification describes a method of determining
 CC whether a compound stimulates a Th1-like response. Th1 cells are a subset
 CC of CD4+ T lymphocyte cells. The method comprises contacting naive
 CC lymphocytes in vitro with a fusion protein comprising at least a fragment
 CC of Hsp, and then detecting the Th1-like response exhibited by the cell
 CC sample. The proteins which may be used in the method of the invention are

CC Hsp65, Hsp40, Hsp10, Hsp60, and Hsp71. The method may be used to identify
 CC compounds that stimulate Th1-like responses in response to microbial
 CC pathogens.
 CC
 XX
 SQ Sequence 245 AA;
 Query Match 98.7%; Score 1149; DB 22; Length 245;
 Best Local Similarity 98.6%; Pred. No. 7.9e-113;
 Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKRNKKFELGEPNLPYYID 60
 DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKRNKKFELGEPNLPYYID 60
 QY 61 GDVKLTQSMALIRYIADKHNMLGSGPKERAEISMLEGAVLDIRGVSRIVASKDEFTLKV 120
 DB 61 GDVKLTQSMALIRYIADKHNMLGSGPKERAEISMLEGAVLDIRGVSRIVASKDEFTLKV 120
 QY 121 DFLSKLPMLKMFEDRLCHKTYLNGDHYTHDPFMYLDVLYMDPCLDAFRLVSEK 180
 DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHYTHDPFMYLDVLYMDPCLDAFRLVSEK 180
 QY 181 KRIFAIPQIDKYLKSSKTIAMPLOGMQATFGGDPHPK 218
 DB 181 KRIFAIPQIDKYLKSSKTIAMPLOGMQATFGGDPHPK 218
 RESULT 8
 AAM74114
 ID AAM74114 standard; peptide: 247 AA.
 XX
 AC AAM74114;
 XX
 DT 04-MAY-1999 (first entry)
 XX
 DE GST/GI transport receptor binding protein fusion sequence.
 XX
 KM Gastro-intestinal transport receptor; binding protein: hSI; HPT1;
 KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KW intestinal peptide-associated transporter; hypertension; diabetes;
 KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KW therapeutic agent delivery; therapy; fusion protein.
 XX
 OS Homo sapiens.
 XX
 PN WO9851325-A2.
 XX
 PD 19-NOV-1998.
 XX
 PF 15-MAY-1998; 98WO-US10088.
 XX
 PR 15-MAY-1997; 97US-0046595.
 XX
 PA (CYTO-) CYTOGEN CORP.
 PA (ELAN-) ELAN CORP PLC.
 XX
 PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin JI;
 PI Mahony DJ, Patterson CA, Singleton J;
 XX
 DR WPI; 1999-009568/01.
 XX
 PT New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimerae and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide
 XX
 PS Example; Page 217; 294pp; English.
 XX
 CC This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (HPT1),

CC hPEPT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.
 XX
 SQ Sequence 247 AA;
 Query Match 98.7%; Score 1149; DB 20; Length 247;
 Best Local Similarity 98.6%; Pred. No. 7.9e-113;
 Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKRNKKFELGEPNLPYYID 60
 DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKRNKKFELGEPNLPYYID 60
 QY 61 GDVKLTQSMALIRYIADKHNMLGSGPKERAEISMLEGAVLDIRGVSRIVASKDEFTLKV 120
 DB 61 GDVKLTQSMALIRYIADKHNMLGSGPKERAEISMLEGAVLDIRGVSRIVASKDEFTLKV 120
 QY 121 DFLSKLPMLKMFEDRLCHKTYLNGDHYTHDPFMYLDVLYMDPCLDAFRLVSEK 180
 DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHYTHDPFMYLDVLYMDPCLDAFRLVSEK 180
 QY 181 KRIFAIPQIDKYLKSSKTIAMPLOGMQATFGGDPHPK 218
 DB 181 KRIFAIPQIDKYLKSSKTIAMPLOGMQATFGGDPHPK 218
 RESULT 9
 AAM74120
 ID AAM74120 standard; peptide: 247 AA.
 XX
 AC AAM74120;
 XX
 DT 04-MAY-1999 (first entry)
 XX
 DE GST/GI transport receptor binding protein fusion sequence.
 XX
 KM Gastro-intestinal transport receptor; binding protein: hSI; HPT1;
 KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KW intestinal peptide-associated transporter; hypertension; diabetes;
 KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KW therapeutic agent delivery; therapy; fusion protein.
 XX
 OS Homo sapiens.
 XX
 PN WO9851325-A2.
 XX
 PD 19-NOV-1998.
 XX
 PF 15-MAY-1998; 98WO-US10088.
 XX
 PR 15-MAY-1997; 97US-0046595.
 XX
 PA (CYTO-) CYTOGEN CORP.
 PA (ELAN-) ELAN CORP PLC.
 XX
 PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin JI;
 PI Mahony DJ, Patterson CA, Singleton J;
 XX
 DR WPI; 1999-009568/01.
 XX
 PT New proteins that bind specifically to receptors in the

XX

KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
intestinal sodium-co-transporter; hypertension; diabetes

KW Intestinal peptide-associated transporter; hypertension; diabetes;

KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KW therapeutic agent delivery; therapy; fusion protein.
 XX Homo sapiens.
 XX W09851325-A2.
 PN 19-NOV-1998.
 PD 19-NOV-1998.
 XX 15-MAY-1998; 98WO-US10088.
 PF 15-MAY-1997; 97US-0046595.
 PR 15-MAY-1997; 97US-0046595.
 PA (CYTO-) CYTOGEN CORP.
 PA (ELAN-) ELAN CORP PLC.
 PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin LJ;
 PI Onahony DJ, Patterson CA, Singleton J;
 DR WPI: 1999-009568/01.
 XX
 XX
 PT New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimaeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide
 XX
 XX Example; Page 215-216; 294pp; English.
 PS
 CC This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (hPIT),
 CC hPPT1, D2H and human sucrose-isomaltase complex (hsi). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.
 CC
 XX
 SO Sequence 248 AA;
 Query Match 98.7%; Score 1149; DB 20; Length 248;
 Best Local Similarity 98.6%; Pred. No. 8e-113;
 Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSPILGVYKIRIGVOPTRLLLEVEEKEEHLHYERDEGDKMNKKFEGLPEPNLPYYID 60
 DB 1 MSPILGVYKIRIGVOPTRLLLEVEEKEEHLHYERDEGDKMNKKFEGLPEPNLPYYID 60
 QY 61 GDVKLTQSMATIRYIADKHNMLGSGPKERAETISMEGAVLDIRYGVSRAYSKDEFTLKV 120
 DB 61 GDVKLTQSMATIRYIADKHNMLGSGPKERAETISMEGAVLDIRYGVSRAYSKDEFTLKV 120
 QY 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHYTHDPFMYDALDVLVYMDPCLDAFPPLVSFK 180
 DB 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHYTHDPFMYDALDVLVYMDPCLDAFPPLVSFK 180
 QY 181 KRIEAIPOIDKYKSKSKTIAMPLOGWQATFGGDDHPPK 218
 DB 181 KRIEAIPOIDKYKSKSKTIAMPLOGWQATFGGDDHPPK 218

ID AAW74113 standard; peptide: 248 AA.
 AC AAW74113;
 XX
 XX
 DT 04-MAY-1999 (first entry)
 DE GST/GI transport receptor binding protein fusion sequence.
 KW Gastro-intestinal transport receptor; binding protein; hsi; hPIT1;
 KW D2H; hPPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KW intestinal peptide-associated transporter; hypertension; diabetes;
 KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KW therapeutic agent delivery; therapy; fusion protein.
 XX
 XX Homo sapiens.
 PN W09851325-A2.
 PD 19-NOV-1998.
 XX 15-MAY-1998; 98WO-US10088.
 PF 15-MAY-1997; 97US-0046595.
 PR 15-MAY-1997; 97US-0046595.
 PA (CYTO-) CYTOGEN CORP.
 PA (ELAN-) ELAN CORP PLC.
 PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin LJ;
 PI Onahony DJ, Patterson CA, Singleton J;
 DR WPI: 1999-009568/01.
 XX
 XX
 PT New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimaeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide
 XX
 XX Example; Page 216; 294pp; English.
 PS
 CC This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (hPIT),
 CC hPPT1, D2H and human sucrose-isomaltase complex (hsi). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.
 CC
 XX
 SO Sequence 248 AA;
 Query Match 98.7%; Score 1149; DB 20; Length 248;
 Best Local Similarity 98.6%; Pred. No. 8e-113;
 Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSPILGVYKIRIGVOPTRLLLEVEEKEEHLHYERDEGDKMNKKFEGLPEPNLPYYID 60
 DB 1 MSPILGVYKIRIGVOPTRLLLEVEEKEEHLHYERDEGDKMNKKFEGLPEPNLPYYID 60
 QY 61 GDVKLTQSMATIRYIADKHNMLGSGPKERAETISMEGAVLDIRYGVSRAYSKDEFTLKV 120
 DB 61 GDVKLTQSMATIRYIADKHNMLGSGPKERAETISMEGAVLDIRYGVSRAYSKDEFTLKV 120

QY 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTVHPDFMLYDALDVLYMDPKLDAFPKLVSPK 180
 DB 121 DFLSKLPEMLKMFEDRLCHKITLNGDHTVHPDFMLYDALDVLYMDPKLDAFPKLVSPK 180
 QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGSDHPK 218
 DB 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGSDHPK 218

RESULT 13

AAW74121
 ID AAW74121 standard; peptide: 248 AA.

AAW74121;

04-MAY-1999 (first entry)

GST/GI transport receptor binding protein fusion sequence.

KW Gastro-intestinal transport receptor; binding protein; hsi; HPT1;
 KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KW intestinal peptide-associated transporter; hypertension; diabetes;
 KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KW therapeutic agent delivery; therapy; fusion protein.

OS Homo sapiens.

XX W09851325-A2.

PD 19-NOV-1998.

PF 15-MAY-1998; 98WO-US10088.

PR 15-MAY-1997; 97US-0046595.

PA (CYTO-) CYTOGEN CORP.

PA (ELAN-) ELAN CORP PLC.

PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin JJ;
 PI Omanony DJ, Patterson CA, Singleton J;

DR WPI; 1999-009568/01.

PT New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimaeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide

XX Example; Page 223; 294pp; English.

CC This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (HPT1),
 CC hPEPT1, D2H and human sucrose-isomaltase complex (hsi). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (i).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.

XX Sequence 248 AA;

Query Match 98.7%; Score 1149; DB 20; Length 248;

Best Local Similarity 98.6%; Pred. No. 8e-113;
 Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLHYERDEGGRANKKFELEGLEPPNLPYYID 60
 DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLHYERDEGGRANKKFELEGLEPPNLPYYID 60
 QY 61 GGVKLTQSMATIRYIADKNNMLGSGPERAETSMLEGAVALDIRYGSRIYASKDEFTLV 120
 DB 61 GGVKLTQSMATIRYIADKNNMLGSGPERAETSMLEGAVALDIRYGSRIYASKDEFTLV 120
 QY 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTVHPDFMLYDALDVLYMDPKLDAFPKLVSPK 180
 DB 121 DFLSKLPEMLKMFEDRLCHKITLNGDHTVHPDFMLYDALDVLYMDPKLDAFPKLVSPK 180
 QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGSDHPK 218
 DB 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGSDHPK 218

RESULT 14

AAW74122
 ID AAW74122 standard; peptide: 248 AA.

AAW74122;

04-MAY-1999 (first entry)

GST/GI transport receptor binding protein fusion sequence.

KW Gastro-intestinal transport receptor; binding protein; hsi; HPT1;
 KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KW intestinal peptide-associated transporter; hypertension; diabetes;
 KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KW therapeutic agent delivery; therapy; fusion protein.

OS Homo sapiens.

XX W09851325-A2.

PD 19-NOV-1998.

PF 15-MAY-1998; 98WO-US10088.

PR 15-MAY-1997; 97US-0046595.

PA (CYTO-) CYTOGEN CORP.

PA (ELAN-) ELAN CORP PLC.

PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin JJ;
 PI Omanony DJ, Patterson CA, Singleton J;

DR WPI; 1999-009568/01.

PT New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimaeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide

XX Example; Page 223; 294pp; English.

CC This sequence represents a fusion protein between
 CC glutathione-S-transferase and a gastro-intestinal transport protein
 CC binding peptide. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (HPT1),
 CC hPEPT1, D2H and human sucrose-isomaltase complex (hsi). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents

CC and antigens. (1) may also provide targeting to the GI tract. Other uses
CC of (1) are: (1) to determine the level of specified receptors in a sample
CC (in a binding assay); and (11) to screen for molecules that bind (1).
CC Immunogenic analogues or derivatives of (1) are used to raise antibodies
CC and in immunosays. The antibodies are used to locate, detect and
CC measure (1), e.g. for imaging, monitoring treatment, tissue analysis
CC etc., also for peptide purification and immobilisation.
XX
SQ Sequence 248 AA:
Query Match 98.7%; Score 1149; DB 20; Length 248;
Best Local Similarity 98.6%; Pred. No. 8e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMKNKFEGLPEPNLPYYID 60
DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMKNKFEGLPEPNLPYYID 60
QY 61 GGVKLTQSMATIRYIADKHNMLGSGPKERAETSMLEGAVALDIRGVSRAYSKDFTLVK 120
DB 61 GGVKLTQSMATIRYIADKHNMLGSGPKERAETSMLEGAVALDIRGVSRAYSKDFTLVK 120
QY 121 DFLSKLPEMLKMFEDRLSKHTYLNQDHYTHPDFMLYDALDVLYMDPKLDAFPKLVSEK 180
DB 121 DFLSKLPEMLKMFEDRLSKHTYLNQDHYTHPDFMLYDALDVLYMDPKLDAFPKLVSEK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGWQATFGGDDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGWQATFGGDDHPK 218
RESULT 15
AAW74100
ID AAW74100 standard; peptide: 248 AA.
XX
AC AAW74100;
XX
DT 04-MAY-1999 (first entry)
XX
DE GST/GI transport receptor binding protein fusion sequence.
XX
KW Gastro-intestinal transport receptor; binding protein; hsi; HPT1;
KW D2H; hhep1; human; GI tract receptor; sucrose-isomaltase complex;
KW intestinal peptide-associated transporter; hypertension; diabetes;
KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
KW therapeutic agent delivery; therapy; fusion protein.
XX
OS Homo sapiens.
XX
PN WO9851325-A2.
XX
PD 19-NOV-1998.
XX
PF 15-MAY-1998; 98WO-US10088.
XX
PR 15-MAY-1997; 97US-0046595.
XX
PA (CYTO-) CYTOGEN CORP.
PA (ELAN-) ELAN CORP PLC.
XX
PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambdin JT;
XX Omahony DJ, Patterson CA, Singleton J;
XX WPI, 1999-009568/01.
XX
XX
XX New proteins that bind specifically to receptors in the
XX gastro-intestinal tract and related nucleic acid - chimaeras and
XX antibodies, used to deliver therapeutic or diagnostic agents to, or
XX through, the gastrointestinal tract, e.g. insulin or leuprolide
XX
XX Example: Page 207; 294pp; English.
XX
XX This sequence represents a fusion protein between

CC glutathione-S-transferase and a gastro-intestinal transport protein
CC binding peptide. The invention relates to purified proteins (1) that bind
CC specifically to at least one of the gastro-intestinal (GI) tract
CC receptors human intestinal peptide-associated transporter (HPT1),
CC hhep1, D2H and human sucrose-isomaltase complex (hsi). (1) provide
CC active transport of therapeutic agents through human and animal GI tissue
CC (into the blood) for in vivo delivery, particularly for treatment or
CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
CC cancer, migraine, or angina pectoris. Specifically they are used to
CC deliver insulin or leuprolide, but many other suitable therapeutic agents
CC are disclosed. Including genes or inhibitory nucleic acid, imaging agents
CC and antigens. (1) may also provide targeting to the GI tract. Other uses
CC of (1) are: (1) to determine the level of specified receptors in a sample
CC (in a binding assay); and (11) to screen for molecules that bind (1)
CC Immunogenic analogues or derivatives of (1) are used to raise antibodies
CC and in immunosays. The antibodies are used to locate, detect and
CC measure (1), e.g. for imaging, monitoring treatment, tissue analysis
CC etc., also for peptide purification and immobilisation.
XX
SQ Sequence 248 AA:
Query Match 98.7%; Score 1149; DB 20; Length 248;
Best Local Similarity 98.6%; Pred. No. 8e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMKNKFEGLPEPNLPYYID 60
DB 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLERDEGDKMKNKFEGLPEPNLPYYID 60
QY 61 GGVKLTQSMATIRYIADKHNMLGSGPKERAETSMLEGAVALDIRGVSRAYSKDFTLVK 120
DB 61 GGVKLTQSMATIRYIADKHNMLGSGPKERAETSMLEGAVALDIRGVSRAYSKDFTLVK 120
QY 121 DFLSKLPEMLKMFEDRLSKHTYLNQDHYTHPDFMLYDALDVLYMDPKLDAFPKLVSEK 180
DB 121 DFLSKLPEMLKMFEDRLSKHTYLNQDHYTHPDFMLYDALDVLYMDPKLDAFPKLVSEK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGWQATFGGDDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGWQATFGGDDHPK 218
Search completed: May 20, 2003, 11:33:55
Job time : 19.4058 secs

Tue May 20 13:19:21 2003

us-10-081-408-5.ra1

Page 1

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:30:54 ; Search time 7.02587 Seconds
(without alignments)
912.939 Million cell updates/sec

Title: US-10-081-408-5
Perfect score: 1164
Sequence: 1 MSPILGYWKIKGLVPTRL.....IAMPLOGMAGATFGGDHPRK 218

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29423922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/Backfilest1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1149	98.7	218	1 US-08-470-837-26	Sequence 26, Appl
2	1149	98.7	218	4 US-08-868-452-26	Sequence 26, Appl
3	1149	98.7	231	4 US-08-129-722A-2	Sequence 2, Appl
4	1149	98.7	232	4 US-08-327-874A-6	Sequence 6, Appl
5	1149	98.7	232	5 PCT-US94-09700-6	Sequence 6, Appl
6	1149	98.7	272	4 US-08-910-820-4	Sequence 4, Appl
7	1149	98.7	272	4 US-08-910-820-6	Sequence 4, Appl
8	1149	98.7	282	4 US-08-910-820-3	Sequence 3, Appl
9	1149	98.7	282	4 US-08-910-820-5	Sequence 3, Appl
10	1149	98.7	307	2 US-08-216-894-6	Sequence 6, Appl
11	1149	98.7	307	4 US-09-115-746-6	Sequence 6, Appl
12	1149	98.7	331	4 US-09-217-228-6	Sequence 6, Appl
13	1149	98.7	352	1 US-08-395-507-1	Sequence 1, Appl
14	1149	98.7	354	2 US-08-216-894-4	Sequence 4, Appl
15	1149	98.7	354	4 US-09-115-746-4	Sequence 4, Appl
16	1149	98.7	362	1 US-08-395-507-2	Sequence 2, Appl
17	1149	98.7	397	4 US-08-327-874A-11	Sequence 11, Appl
18	1149	98.7	397	5 PCT-US94-09700-11	Sequence 11, Appl
19	1149	98.7	412	4 US-09-366-009-34	Sequence 34, Appl
20	1149	98.7	422	4 US-09-217-228-7	Sequence 7, Appl
21	1149	98.7	426	3 US-08-737-248-4	Sequence 4, Appl
22	1149	98.7	435	5 PCT-US95-04439-1	Sequence 1, Appl
23	1149	98.7	439	4 US-08-506-296B-67	Sequence 67, Appl
24	1149	98.7	442	4 US-08-506-296B-70	Sequence 70, Appl
25	1149	98.7	443	4 US-08-506-296B-76	Sequence 76, Appl
26	1149	98.7	447	4 US-08-506-296B-73	Sequence 73, Appl
27	1149	98.7	472	2 US-08-216-894-10	Sequence 10, Appl

28	1149	98.7	472	4 US-09-115-746-10	Sequence 10, Appl
29	1149	98.7	514	4 US-08-974-549A-605	Sequence 605, App
30	1149	98.7	515	4 US-08-974-549A-606	Sequence 604, App
31	1149	98.7	517	4 US-08-974-549A-606	Sequence 606, App
32	1149	98.7	530	4 US-08-974-549A-603	Sequence 603, App
33	1149	98.7	536	3 US-08-974-180-20	Sequence 602, App
34	1149	98.7	538	4 US-08-506-296B-69	Sequence 69, Appl
35	1149	98.7	542	4 US-08-506-296B-75	Sequence 75, Appl
36	1149	98.7	547	4 US-08-506-296B-72	Sequence 72, Appl
37	1149	98.7	547	4 US-08-506-296B-66	Sequence 66, Appl
38	1149	98.7	559	2 US-08-216-894-2	Sequence 2, Appl
39	1149	98.7	564	4 US-09-115-746-2	Sequence 2, Appl
40	1149	98.7	579	2 US-08-864-234-11	Sequence 11, Appl
41	1149	98.7	632	4 US-08-506-296B-74	Sequence 74, Appl
42	1149	98.7	633	4 US-08-506-296B-71	Sequence 71, Appl
43	1149	98.7	643	4 US-08-216-894-8	Sequence 8, Appl
44	1149	98.7	643	2 US-09-115-746-8	Sequence 8, Appl
45	1149	98.7	643	4 US-09-115-746-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-08-470-837-26
; Sequence 26, Application US/08470837
; Patent No. 5800811
; GENERAL INFORMATION:
; APPLICANT: Nimmi, Marcel E.
; APPLICANT: Hall, Frederick L.
; APPLICANT: Tuan, Tai-Lan
; APPLICANT: Wu, Lingtao
; APPLICANT: Cheung, David T.
; TITLE OF INVENTION: Transforming Growth Factor B Fusion
; TITLE OF INVENTION: and
; TITLE OF INVENTION: Their Use in Wound Healing
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Merchant & Gould
; STREET: 11150 Santa Monica Boulevard, Suite 400
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90025-3395
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,837
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharp, Janice A.
; REGISTRATION NUMBER: 34,051
; REFERENCE/DOCKET NUMBER: 30630-1US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 310-445-1140
; TELEFAX: 310-445-9031
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 218 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-470-837-26
; Query Match 98.7%; Score 1149; DB 1; Length 218;
; Best Local Similarity 98.6%; Pred. No. 6.8e-114;
; Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
; Oy 1 MSPILGYWKIKGLVPTRLLEYLEEKYEHLHYRDECDKRNKKEFLGLEFPNLPYYID 60

Db 1 MSPIIGYWKIKGLVOPTRLLLEYLEEKYEHLERDEGDKMKNKFFELGLEFPNLPYYID 60
Qy 61 GDVKLTQSMATIRYADKHNMLGSGPKERAETSMLEGAVALDIRGVSRVSIASKDEETLKV 120
Db 61 GDVKLTQSMATIRYADKHNMLGSGPKERAETSMLEGAVALDIRGVSRVSIASKDEETLKV 120
Qy 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPCLDAFPKLVSEK 180
Db 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPCLDAFPKLVSEK 180
Qy 181 KRIEAIPOIDKYLKSSKYIAMPLOGWQATFGGDDHPPK 218
Db 181 KRIEAIPOIDKYLKSSKYIAMPLOGWQATFGGDDHPPK 218

RESULT 2

US-08-868-452-26
; Sequence 26, Application US/08868452C
; Patent No. 6352972
; GENERAL INFORMATION:
; APPLICANT: Marcel E. Nimni
; APPLICANT: Frederick L. Hall
; APPLICANT: Lingtao Wu
; APPLICANT: Bo Han
; APPLICANT: Edwin Shors
; TITLE OF INVENTION: BONE MORPHOGENETIC PROTEINS AND THEIR
; TITLE OF INVENTION: USE IN BONE GROWTH
; FILE REFERENCE: 17972-11
; CURRENT APPLICATION NUMBER: US/08/868.452C
; CURRENT FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 26
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-08-868-452-26

Query Match 98.7%; Score 1149; DB 4; Length 218;
Best Local Similarity 98.6%; Pred. No. 6.8e-114;

Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MSPIIGYWKIKGLVOPTRLLLEYLEEKYEHLERDEGDKMKNKFFELGLEFPNLPYYID 60
Db 1 MSPIIGYWKIKGLVOPTRLLLEYLEEKYEHLERDEGDKMKNKFFELGLEFPNLPYYID 60
Qy 61 GDVKLTQSMATIRYADKHNMLGSGPKERAETSMLEGAVALDIRGVSRVSIASKDEETLKV 120
Db 61 GDVKLTQSMATIRYADKHNMLGSGPKERAETSMLEGAVALDIRGVSRVSIASKDEETLKV 120
Qy 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPCLDAFPKLVSEK 180
Db 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPCLDAFPKLVSEK 180
Qy 181 KRIEAIPOIDKYLKSSKYIAMPLOGWQATFGGDDHPPK 218
Db 181 KRIEAIPOIDKYLKSSKYIAMPLOGWQATFGGDDHPPK 218

RESULT 3

US-08-129-722A-2
; Sequence 2, Application US/08129722A
; Patent No. 6303369
; GENERAL INFORMATION:
; APPLICANT: Spana, Carl
; APPLICANT: Fargnoli, Joseph
; APPLICANT: Bolen, Joseph B.
; TITLE OF INVENTION: PROTEIN EXPRESSION SYSTEM
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burton Rodney
; STREET: P.O. Box 4000

CITY: Princeton
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08543-4000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/129,722A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gaul, Timothy J.
REGISTRATION NUMBER: 33,111
REFERENCE/DOCKET NUMBER: DC25
TELEPHONE: (609) 252-5901
TELEFAX: (609) 252-4526
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 231 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-129-722A-2

Query Match 98.7%; Score 1149; DB 4; Length 231;
Best Local Similarity 98.6%; Pred. No. 7.4e-114;

Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MSPIIGYWKIKGLVOPTRLLLEYLEEKYEHLERDEGDKMKNKFFELGLEFPNLPYYID 60
Db 1 MSPIIGYWKIKGLVOPTRLLLEYLEEKYEHLERDEGDKMKNKFFELGLEFPNLPYYID 60
Qy 61 GDVKLTQSMATIRYADKHNMLGSGPKERAETSMLEGAVALDIRGVSRVSIASKDEETLKV 120
Db 61 GDVKLTQSMATIRYADKHNMLGSGPKERAETSMLEGAVALDIRGVSRVSIASKDEETLKV 120
Qy 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPCLDAFPKLVSEK 180
Db 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPCLDAFPKLVSEK 180
Qy 181 KRIEAIPOIDKYLKSSKYIAMPLOGWQATFGGDDHPPK 218
Db 181 KRIEAIPOIDKYLKSSKYIAMPLOGWQATFGGDDHPPK 218

RESULT 4

US-08-327-874A-6
; Sequence 6, Application US/08327874A
; Patent No. 6372249
; GENERAL INFORMATION:
; APPLICANT: BAYLOR COLLEGE OF MEDICINE
; APPLICANT: SMITH, JAMES R.
; APPLICANT: DRUTZ, DAVID J.
; APPLICANT: WILSON, DEBORAH R.
; APPLICANT: ZUMSTEIN, LOUIS A.
; TITLE OF INVENTION: SENSITIVE CELL DERIVED INHIBITORS OF
; TITLE OF INVENTION: DNA SYNTHESIS
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROGERS & WELLS
; STREET: 200 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10166
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/327,874A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/808,523
FILING DATE: 16-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/970,462
FILING DATE: 02-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/113,372
FILING DATE: 30-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,564
FILING DATE: 17-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/203,535
FILING DATE: 25-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/229,420
FILING DATE: 15-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/274,535
FILING DATE: 13-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT US94/09700
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/327,874
FILING DATE: 24-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: NORTON, GERRARD P.
REGISTRATION NUMBER: 36,621
REFERENCE/DOCKET NUMBER: 3634-8-CIP10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 878-3148
TELEFAX: (212) 878-8375
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 232 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Schistosoma japonicum
IMMEDIATE SOURCE:
CLONE: GST
US-08-327-874A-6

Query Match 98.7%; Score 1149; DB 4; Length 232;
Best Local Similarity 98.6%; Pred. No. 7.4e-114;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGYKIKGLVQPTRLLEYLEEKYEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYKIKGLVQPTRLLEYLEEKYEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
QY 61 GDVKLTOSMAIIRYIADRNHMLGSPKRAEISMLEGAVALDIRYVSRIAYSKDEFTLKV 120
DB 61 GDVKLTOSMAIIRYIADRNHMLGSPKRAEISMLEGAVALDIRYVSRIAYSKDEFTLKV 120
QY 121 DFLSLPMLKMFEDRLSHKTYTLNGDHYTHPDMLYDALDVLVLYNDPCLDAFPLVSEFK 180
DB 121 DFLSLPMLKMFEDRLSHKTYTLNGDHYTHPDMLYDALDVLVLYNDPCLDAFPLVSEFK 180
QY 181 KRIEAIPOIDKYLKSSXYIAMPLOGWQATFGGDHPRK 218
DB 181 KRIEAIPOIDKYLKSSXYIAMPLOGWQATFGGDHPRK 218

RESULT 5
PCT-US94-09700-6
Sequence 6, Application PC/TUS9409700
GENERAL INFORMATION:
APPLICANT: BAYLOR COLLEGE OF MEDICINE
APPLICANT: SMITH, JAMES R.
TITLE OF INVENTION: SENSICENT CELL DERIVED INHIBITORS OF
TITLE OF INVENTION: DNA SYNTHESIS
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOWREY & SIMON
STREET: 1299 PENNSYLVANIA AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09700
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/808,523
FILING DATE: 16-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/970,462
FILING DATE: 02-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/113,372
FILING DATE: 30-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/153,564
FILING DATE: 17-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/203,535
FILING DATE: 25-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/229,420
FILING DATE: 15-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/274,535
FILING DATE: 13-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: AUERBACH, JEFFREY I.
REGISTRATION NUMBER: 32,680
REFERENCE/DOCKET NUMBER: 225-102-CIP7-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 383-7451
TELEFAX: (202) 383-6610
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 232 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Schistosoma japonicum
IMMEDIATE SOURCE:
CLONE: GST
PCT-US94-09700-6

Query Match 98.7%; Score 1149; DB 5; Length 232;
Best Local Similarity 98.6%; Pred. No. 7.4e-114;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGYKIKGLVQPTRLLEYLEEKYEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYKIKGLVQPTRLLEYLEEKYEHLHYERDEGDKRNKKFELGLEFPNLPYYID 60

Sequence 3, Application US/08910820
Patent No. 6258579
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
APPLICANT: Zhu, Hengyi
APPLICANT: Barbosa, Miguel
APPLICANT: Li, Gian
APPLICANT: Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
NUMBER OF SEQUENCES: 25
TITLE OF INVENTION: COMPLEX AND METHODS OF USE THEREFOR
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-910-820-3

Query Match 98.7%; Score 1149; DB 4; Length 282;
Best Local Similarity 98.6%; Pred. No. 9.8e-114;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGWYKIKGLVQPTLLLEYLEEKEEHLERDEGDKWNKKFELGLEFPNLPYYID 60
DB 1 MSPILGWYKIKGLVQPTLLLEYLEEKEEHLERDEGDKWNKKFELGLEFPNLPYYID 60
QY 61 GDVKLTQSMALIRYIADHNHMGSPKERAETSMLEGAVLDIRYGSRIAYSKDEETLKV 120
DB 61 GDVKLTQSMALIRYIADHNHMGSPKERAETSMLEGAVLDIRYGSRIAYSKDEETLKV 120
QY 121 DFLSKLPENLKMFEEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPCLDAFPKLVYFK 180
DB 121 DFLSKLPENLKMFEEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPCLDAFPKLVYFK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGWQATFGGDDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGWQATFGGDDHPK 218

RESULT 9
US-08-910-820-5
Sequence 5, Application US/08910820
Patent No. 6258579
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
APPLICANT: Zhu, Hengyi
APPLICANT: Barbosa, Miguel
APPLICANT: Li, Gian
APPLICANT: Murray, Brian W.

TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-910-820-5

Query Match 98.7%; Score 1149; DB 4; Length 282;
Best Local Similarity 98.6%; Pred. No. 9.8e-114;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGWYKIKGLVQPTLLLEYLEEKEEHLERDEGDKWNKKFELGLEFPNLPYYID 60
DB 1 MSPILGWYKIKGLVQPTLLLEYLEEKEEHLERDEGDKWNKKFELGLEFPNLPYYID 60
QY 61 GDVKLTQSMALIRYIADHNHMGSPKERAETSMLEGAVLDIRYGSRIAYSKDEETLKV 120
DB 61 GDVKLTQSMALIRYIADHNHMGSPKERAETSMLEGAVLDIRYGSRIAYSKDEETLKV 120
QY 121 DFLSKLPENLKMFEEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPCLDAFPKLVYFK 180
DB 121 DFLSKLPENLKMFEEDRLSHKTYLNGDHYTHPDFMLYDALDVLYMDPCLDAFPKLVYFK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGWQATFGGDDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGWQATFGGDDHPK 218

RESULT 10
US-08-216-894-6
Sequence 6, Application US/08216894
Patent No. 5876734
GENERAL INFORMATION:
APPLICANT: Kirchhoff, Louis V.
APPLICANT: Otsu, Kelko
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/216,894
FILING DATE: 24-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 85326/102/DRLO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 307 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-216-894-6

Query Match 98.7%; Score 1149; DB 2; Length 307;
Best Local Similarity 98.6%; Pred. No. 1,1e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 MSPILGYKIKGLVQPTLLLEYLEEKYEENLYERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYKIKGLVQPTLLLEYLEEKYEENLYERDEGDKRNKKFELGLEFPNLPYYID 60
OY 61 GVKLTQSMATIRYIADKHNMLGSGPKERAETISMEGAVLDIRYGVSRIVASKDFETLKV 120
DB 61 GVKLTQSMATIRYIADKHNMLGSGPKERAETISMEGAVLDIRYGVSRIVASKDFETLKV 120
OY 121 DFLSKLPEMLKMFEDRLSHKTYTLNGDHYTHDPFMYDALDVLYMDPCLDAFPKLYSEFK 180
DB 121 DFLSKLPEMLKMFEDRLCHKTYTLNGDHYTHDPFMYDALDVLYMDPCLDAFPKLYSEFK 180
OY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFFGGGDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFFGGGDHPK 218

RESULT 11

US-09-115-746-6
Sequence 6, Application US/09115746
Patent No. 6228601
GENERAL INFORMATION:
APPLICANT: Kirchoff, Louis V.
APPLICANT: Otsu, Keiko
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
TITLE OF INVENTION: WITH TRYPAPOSOMA CRUZI
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/115,746
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/216,894
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 85326/102/DRLO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 307 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-115-746-6

Query Match 98.7%; Score 1149; DB 4; Length 307;
Best Local Similarity 98.6%; Pred. No. 1,1e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 MSPILGYKIKGLVQPTLLLEYLEEKYEENLYERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYKIKGLVQPTLLLEYLEEKYEENLYERDEGDKRNKKFELGLEFPNLPYYID 60
OY 61 GVKLTQSMATIRYIADKHNMLGSGPKERAETISMEGAVLDIRYGVSRIVASKDFETLKV 120
DB 61 GVKLTQSMATIRYIADKHNMLGSGPKERAETISMEGAVLDIRYGVSRIVASKDFETLKV 120
OY 121 DFLSKLPEMLKMFEDRLSHKTYTLNGDHYTHDPFMYDALDVLYMDPCLDAFPKLYSEFK 180
DB 121 DFLSKLPEMLKMFEDRLCHKTYTLNGDHYTHDPFMYDALDVLYMDPCLDAFPKLYSEFK 180
OY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFFGGGDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFFGGGDHPK 218

RESULT 12

US-09-217-228-6
Sequence 6, Application US/09217228
Patent No. 6323178
GENERAL INFORMATION:
APPLICANT: Butler, Jon P.
APPLICANT: Hale, John E.
APPLICANT: Heath Jr., William F.
APPLICANT: Schoner, Brigitte E.
APPLICANT: Helman, Mark L.
APPLICANT: Becker, Gerald W.
APPLICANT: Varshavsky, Alexander D.
TITLE OF INVENTION: Beta-1ipotrofin and Uses Thereof
FILE REFERENCE: X-12139
CURRENT APPLICATION NUMBER: US/09/217,228
NUMBER OF SEQ ID NOS: 35
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 331
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: GST fusion
US-09-217-228-6

Query Match 98.7%; Score 1149; DB 4; Length 331;
Best Local Similarity 98.6%; Pred. No. 1,2e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 MSPILGYKIKGLVQPTLLLEYLEEKYEENLYERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYKIKGLVQPTLLLEYLEEKYEENLYERDEGDKRNKKFELGLEFPNLPYYID 60
OY 61 GVKLTQSMATIRYIADKHNMLGSGPKERAETISMEGAVLDIRYGVSRIVASKDFETLKV 120
DB 61 GVKLTQSMATIRYIADKHNMLGSGPKERAETISMEGAVLDIRYGVSRIVASKDFETLKV 120

OY 121 DFLSKLPMLKMFEDRLSHKTYTLNGDHYTHDPFMYDALDVLYMDPNCCLDAFPKLYSEK 180
DB 121 DFLSKLPMLKMFEDRLCHKTYTLNGDHYTHDPFMYDALDVLYMDPNCCLDAFPKLYSEK 180
OY 181 KRIEAIPOIDKYLKSSKTYIAMPLOGWQATFGGDPHPK 218
DB 181 KRIEAIPOIDKYLKSSKTYIAMPLOGWQATFGGDPHPK 218

RESULT 13

US-08-395-507-1
; Sequence 1, Application US/08395507
; Patent No. 5578456
; GENERAL INFORMATION:
; APPLICANT: Fujimura, Katsuya
; APPLICANT: Ueno, Eiichi
; APPLICANT: Fujii, No. 5578456uyuk1
; APPLICANT: Okada, Masahisa
; TITLE OF INVENTION: Anti-tetraponema Pallidum Antibody
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pollock, Vande Sande & Priddy
; STREET: 1990 M Street, NW, Suite 800
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/395,507
; FILING DATE: 27-FEB-1995
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP Hei-6-54672
; FILING DATE: 28-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Amernick, Burton A.
; REGISTRATION NUMBER: 24,852
; REFERENCE/DOCKET NUMBER: 1581/007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 331-7111
; TELEFAX: (202) 293-6229
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 352 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: not relevant
; MOLECULE TYPE: Protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: rabbit
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..352
; OTHER INFORMATION: /note= "G15 Antigen"
US-08-395-507-1

Query Match 98.7%; Score 1149; DB 1; Length 352;
Best Local Similarity 98.6%; Pred. No. 1.3e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 MSPILGWYKIRKGLVPTLLEYLEEKYEHLHYERDEGDKWRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGWYKIRKGLVPTLLEYLEEKYEHLHYERDEGDKWRNKKFELGLEFPNLPYYID 60
OY 61 GDVKLTQSMATIRYIADKHNHMLGSGPKRAEISMLEGAVLDIRGVSRIVASKDEFTLVK 120

DB 61 GDVKLTQSMATIRYIADKHNHMLGSGPKRAEISMLEGAVLDIRGVSRIVASKDEFTLVK 120
OY 121 DFLSKLPMLKMFEDRLSHKTYTLNGDHYTHDPFMYDALDVLYMDPNCCLDAFPKLYSEK 180
DB 121 DFLSKLPMLKMFEDRLCHKTYTLNGDHYTHDPFMYDALDVLYMDPNCCLDAFPKLYSEK 180
OY 181 KRIEAIPOIDKYLKSSKTYIAMPLOGWQATFGGDPHPK 218
DB 181 KRIEAIPOIDKYLKSSKTYIAMPLOGWQATFGGDPHPK 218

RESULT 14

US-08-216-894-4
; Sequence 4, Application US/08216894
; Patent No. 5876734
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,894
; FILING DATE: 24-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRL0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 354 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-216-894-4

Query Match 98.7%; Score 1149; DB 2; Length 354;
Best Local Similarity 98.6%; Pred. No. 1.4e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 MSPILGWYKIRKGLVPTLLEYLEEKYEHLHYERDEGDKWRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGWYKIRKGLVPTLLEYLEEKYEHLHYERDEGDKWRNKKFELGLEFPNLPYYID 60
OY 61 GDVKLTQSMATIRYIADKHNHMLGSGPKRAEISMLEGAVLDIRGVSRIVASKDEFTLVK 120
DB 61 GDVKLTQSMATIRYIADKHNHMLGSGPKRAEISMLEGAVLDIRGVSRIVASKDEFTLVK 120
OY 121 DFLSKLPMLKMFEDRLSHKTYTLNGDHYTHDPFMYDALDVLYMDPNCCLDAFPKLYSEK 180
DB 121 DFLSKLPMLKMFEDRLCHKTYTLNGDHYTHDPFMYDALDVLYMDPNCCLDAFPKLYSEK 180
OY 181 KRIEAIPOIDKYLKSSKTYIAMPLOGWQATFGGDPHPK 218
DB 181 KRIEAIPOIDKYLKSSKTYIAMPLOGWQATFGGDPHPK 218

RESULT 15
US-09-115-746-4
: Sequence 4, Application US/09115746
: Patent No. 6228601
: GENERAL INFORMATION:
: APPLICANT: Kirchhoff, Louis V.
: APPLICANT: Otsu, Keiko
: TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
: TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W., Suite 500
: CITY: Washington, D.C.
: COUNTRY: USA
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/115,746
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/216,894
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 85326/102/DBLO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)672-5300
: TELEFAX: (202)672-5399
: TELEX: 904136
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 354 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-115-746-4

Query Match 98.7%; Score 1149; DB 4; Length 354;
Best Local Similarity 98.6%; Pred. No. 1,4e-113;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEHLRYERDEGDKWRNKKFELGLEFPNLPYYID 60
Db 1 MSPILGYWKIKGLVOPTRLLLEYLEEKYEHLRYERDEGDKWRNKKFELGLEFPNLPYYID 60
QY 61 GDVKLTQSMATIRYIADKHNNLGGSPKRAEISMLEGAVLDIRYGSRIAYSKDPETLKV 120
Db 61 GDVKLTQSMATIRYIADKHNNLGGSPKRAEISMLEGAVLDIRYGSRIAYSKDPETLKV 120
QY 121 DFLSKLPMLKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVVLYMDPMLDAFPPLYSFK 180
Db 121 DFLSKLPMLKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVVLYMDPMLDAFPPLYSFK 180
QY 181 KRIEAIPOIDKYLYKSSKIYAMPLOGWQATFGGSDHPPK 218
Db 181 KRIEAIPOIDKYLYKSSKIYAMPLOGWQATFGGSDHPPK 218

Search completed: May 20, 2003, 11:40:57
Job time : 8.02387 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:37:59 ; Search time 34.7335 Seconds

(Without alignments)
622.443 Million cell updates/sec

Title: US-10-081-408-5

Perfect score: 1164

Sequence: 1 MSPILGYWKIKGLVQPTRL.....IAMPLOQWQATFGGDDHPK 218

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PTI_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/PTI_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1164	100.0	218	US-10-081-408-5	Sequence 5, Appli
2	1159	99.6	998	US-10-081-408-20	Sequence 20, Appli
3	1149	98.7	218	US-10-081-408-4	Sequence 4, Appli
4	1149	98.7	239	US-09-823-153-7	Sequence 7, Appli
5	1149	98.7	245	US-10-267-311-23	Sequence 23, Appli
6	1149	98.7	272	US-09-844-988-4	Sequence 4, Appli
7	1149	98.7	272	US-09-844-988-6	Sequence 6, Appli
8	1149	98.7	272	US-09-844-988-4	Sequence 4, Appli
9	1149	98.7	272	US-09-844-988-3	Sequence 3, Appli
10	1149	98.7	282	US-09-844-988-3	Sequence 3, Appli
11	1149	98.7	282	US-09-844-988-5	Sequence 5, Appli
12	1149	98.7	282	US-09-844-988-3	Sequence 3, Appli
13	1149	98.7	282	US-09-844-988-3	Sequence 3, Appli
14	1149	98.7	298	US-09-910-600-26	Sequence 26, Appli
15	1149	98.7	324	US-10-267-311-25	Sequence 25, Appli
16	1149	98.7	331	US-09-824-438-6	Sequence 6, Appli
17	1149	98.7	348	US-09-910-600-22	Sequence 22, Appli
18	1149	98.7	348	US-09-910-600-23	Sequence 23, Appli
19	1149	98.7	348	US-09-910-600-24	Sequence 24, Appli

20	1149	98.7	348	US-09-910-600-25	Sequence 25, Appli
21	1149	98.7	354	US-09-823-153-8	Sequence 8, Appli
22	1149	98.7	364	US-09-788-268-14	Sequence 14, Appli
23	1149	98.7	394	US-09-980-578-4	Sequence 4, Appli
24	1149	98.7	401	US-10-120-866-1	Sequence 1, Appli
25	1149	98.7	412	US-09-775-964-34	Sequence 34, Appli
26	1149	98.7	422	US-09-824-438-7	Sequence 7, Appli
27	1149	98.7	1140	US-09-950-634-4	Sequence 4, Appli
28	1146	98.5	649	US-10-174-784-9	Sequence 9, Appli
29	483.5	41.5	229	US-10-102-806-554	Sequence 554, App
30	196.5	16.9	222	US-09-784-739-4	Sequence 4, Appli
31	196.5	16.9	230	US-09-925-301-1318	Sequence 1318, Ap
32	179.5	15.4	222	US-09-784-739-3	Sequence 3, Appli
33	169.5	14.6	203	US-09-847-208-48	Sequence 48, Appli
34	169.5	14.6	222	US-09-784-739-5	Sequence 5, Appli
35	161.5	13.9	222	US-09-784-739-1	Sequence 1, Appli
36	104	8.9	240	US-09-964-889-23	Sequence 23, Appli
37	103	8.8	233	US-09-765-213A-2	Sequence 2, Appli
38	96	8.2	233	US-09-765-213A-4	Sequence 4, Appli
39	95	8.2	233	US-09-765-213A-6	Sequence 6, Appli
40	90	7.7	379	US-10-253-007-46	Sequence 46, Appli
41	89	7.6	263	US-09-769-787-68	Sequence 68, Appli
42	88.5	7.6	359	US-09-738-636-4895	Sequence 4895, Ap
43	86	7.4	1668	US-09-815-242-5654	Sequence 5654, Ap
44	86	7.4	2397	US-09-815-242-12265	Sequence 12265, A
45	85.5	7.3	241	US-09-854-133-204	Sequence 204, App

ALIGNMENTS

RESULT 1
US-10-081-408-5
Sequence 5, Application US/10081408
Patent No. US20020160482A1
GENERAL INFORMATION:
APPLICANT: Abirams, n, Lars
TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
FILE REFERENCE: 13425-053001
CURRENT APPLICATION NUMBER: US/10/081,408
CURRENT FILING DATE: 2002-02-21
PRIOR APPLICATION NUMBER: SE 0100625-3
PRIOR FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: US 60/272,247
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 20
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 218
TYPE: PRT
ORGANISM: Schistosoma japonicum
US-10-081-408-5

Query Match	100.0%	Score 1164;	DB 9;	Length 218;
Best Local Similarity	100.0%	Pred. No. 3.4e-106;		
Matches 218;	Conservative	0;	Mismatches 0;	Indels 0;
Gaps 0;				
QY	1	MSPILGYWKIKGLVQPTRLLEYLEEKYEHENYENDEGKRNKKFELGLEPNNLPYYTD	60	
DB	1	MSPILGYWKIKGLVQPTRLLEYLEEKYEHENYENDEGKRNKKFELGLEPNNLPYYTD	60	
QY	61	GVVKLTQSAAIRYIADKNMNLGSPKRAEISMLEGAVLDIRYGSRIAYSKDEFTLKV	120	
DB	61	GVVKLTQSAAIRYIADKNMNLGSPKRAEISMLEGAVLDIRYGSRIAYSKDEFTLKV	120	
QY	121	DFLSKLPKMLKXFEEDRLSHKTYLNGDHYTHPDMYDALDVLVYKDPKCLAFPKLVSEK	180	
DB	121	DFLSKLPKMLKXFEEDRLSHKTYLNGDHYTHPDMYDALDVLVYKDPKCLAFPKLVSEK	180	
QY	181	KRIEAIPOIDKYLKSSKYIAMPLOQWQATFGGDDHPK 218		
DB	181	KRIEAIPOIDKYLKSSKYIAMPLOQWQATFGGDDHPK 218		

RESULT 2
US-10-081-408-20
; Sequence 20, Application US/10081408
; Patent No. US20020160482A1
; GENERAL INFORMATION:
; APPLICANT: Abrahams, n, Lars
; TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
; FILE REFERENCE: 13425-053001
; CURRENT APPLICATION NUMBER: US/10/081,408
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: SE 0100625-3
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 60/272,247
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 998
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Recombinant construct
US-10-081-408-20

Query Match 99.6%; Score 1159; DB 9; Length 998;
Best Local Similarity 100.0%; Pred. No. 7,7e-105;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SPILGYWKIKGLVOPTRLLLEYLEKEYEENLYERDEGDKRNKKFELGLEFPNLPYYIDG 61
DB 38 SPILGYWKIKGLVOPTRLLLEYLEKEYEENLYERDEGDKRNKKFELGLEFPNLPYYIDG 97
QY 62 DYKLTQSMAIIRYIADKHNMLGSGPKERAISMLEGAVLDIRYGVSRIVASKEDETLKY 121
DB 98 DYKLTQSMAIIRYIADKHNMLGSGPKERAISMLEGAVLDIRYGVSRIVASKEDETLKY 157
QY 122 FLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLVYMPDCLDAFPKLVSEFK 181
DB 158 FLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLVYMPDCLDAFPKLVSEFK 217
QY 182 RIEAIPOIDIKYLKSSKYIAMPLOGQATFGGDDHPPK 218
DB 218 RIEAIPOIDIKYLKSSKYIAMPLOGQATFGGDDHPPK 254

RESULT 3
US-10-081-408-4
; Sequence 4, Application US/10081408
; Patent No. US20020160482A1
; GENERAL INFORMATION:
; APPLICANT: Abrahams, n, Lars
; APPLICANT: Nilsson, Joakim
; TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
; FILE REFERENCE: 13425-053001
; CURRENT APPLICATION NUMBER: US/10/081,408
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: SE 0100625-3
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: US 60/272,247
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Schistosoma japonicum
US-10-081-408-4

Query Match 98.7%; Score 1149; DB 9; Length 218;
Best Local Similarity 98.6%; Pred. No. 9,9e-105;

Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVOPTRLLLEYLEKEYEENLYERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYWKIKGLVOPTRLLLEYLEKEYEENLYERDEGDKRNKKFELGLEFPNLPYYID 60
QY 61 GVKLTQSMAIIRYIADKHNMLGSGPKERAISMLEGAVLDIRYGVSRIVASKEDETLKY 120
DB 61 GVKLTQSMAIIRYIADKHNMLGSGPKERAISMLEGAVLDIRYGVSRIVASKEDETLKY 120
QY 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLVYMPDCLDAFPKLVSEFK 180
DB 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLVYMPDCLDAFPKLVSEFK 180
QY 181 KRIEAIPOIDIKYLKSSKYIAMPLOGQATFGGDDHPPK 218
DB 181 KRIEAIPOIDIKYLKSSKYIAMPLOGQATFGGDDHPPK 218

RESULT 4
US-09-823-153-7
; Sequence 7, Application US/09823153
; Patent No. US20020025540A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; APPLICANT: Roberts, Susan
; APPLICANT: Pak, Roger
; APPLICANT: Lewis, Martin
; APPLICANT: Smith, David
; APPLICANT: Hendrick, Joseph
; APPLICANT: Vinitsky, Alexander
; TITLE OF INVENTION: ISOLATION OF FUNCTIONALLY ACTIVE GAMMA-SECRETASE PROTEIN COMPL
; FILE REFERENCE: D0004
; CURRENT APPLICATION NUMBER: US/09/823,153
; CURRENT FILING DATE: 2001-07-02
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Description of Artificial Sequence: PSI PEPTIDE
US-09-823-153-7

Query Match 98.7%; Score 1149; DB 10; Length 239;
Best Local Similarity 98.6%; Pred. No. 1,1e-104;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVOPTRLLLEYLEKEYEENLYERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYWKIKGLVOPTRLLLEYLEKEYEENLYERDEGDKRNKKFELGLEFPNLPYYID 60
QY 61 GVKLTQSMAIIRYIADKHNMLGSGPKERAISMLEGAVLDIRYGVSRIVASKEDETLKY 120
DB 61 GVKLTQSMAIIRYIADKHNMLGSGPKERAISMLEGAVLDIRYGVSRIVASKEDETLKY 120
QY 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLVYMPDCLDAFPKLVSEFK 180
DB 121 DFLSKLPEMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLVYMPDCLDAFPKLVSEFK 180
QY 181 KRIEAIPOIDIKYLKSSKYIAMPLOGQATFGGDDHPPK 218
DB 181 KRIEAIPOIDIKYLKSSKYIAMPLOGQATFGGDDHPPK 218

RESULT 5
US-10-267-311-23
; Sequence 23, Application US/10267311
; Publication No. US20030050469A1
; GENERAL INFORMATION:

APPLICANT: Siegel, Marvin
APPLICANT: Chu, N. Randall
TITLE OF INVENTION: INDUCTION OF A TH1-LIKE RESPONSE IN VITRO
FILE REFERENCE: 12071/002001
CURRENT APPLICATION NUMBER: US/10/267,311
CURRENT FILING DATE: 2002-10-09
PRIOR APPLICATION NUMBER: US/09/613,303
PRIOR FILING DATE: 2000-07-10
PRIOR APPLICATION NUMBER: US 60/143,757
PRIOR FILING DATE: 1999-07-08
NUMBER OF SEQ ID NOS: 55
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23
LENGTH: 245
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: fusion sequence
US-10-267-311-23

Query Match 98.7%; Score 1149; DB 9; Length 245;
Best Local Similarity 98.6%; Pred. No. 1.1e-104;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGKIKIGVOPTRLLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGKIKIGVOPTRLLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYYID 60
QY 61 GDVLTOSMAITRIYADHNMLGSPKERAISMLEGAVLDIRYVSRIATSKDEETLKV 120
DB 61 GDVLTOSMAITRIYADHNMLGSPKERAISMLEGAVLDIRYVSRIATSKDEETLKV 120
QY 121 DFLSKLPMLKMFEDRLSHKTYTLNGDHYTHDPFMLYDALDVLYNDPMLCLDAFPPLVSEFK 180
DB 121 DFLSKLPMLKMFEDRLSHKTYTLNGDHYTHDPFMLYDALDVLYNDPMLCLDAFPPLVSEFK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFFGGGDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFFGGGDHPK 218

RESULT 6
US-09-844-988-4
Sequence 4, Application US/09844988
Patent No. US20020158764A1
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
Zhu, Hengyi
Barbosa, Miguel
Li, Gfan
Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,988
FILING DATE: 26-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/910,820

FILING DATE: 1997-08-13
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-844-988-4

Query Match 98.7%; Score 1149; DB 9; Length 272;
Best Local Similarity 98.6%; Pred. No. 1.3e-104;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGKIKIGVOPTRLLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGKIKIGVOPTRLLLEYLEEKYEHLERDEGDKRNKKFELGLEFPNLPYYID 60
QY 61 GDVLTOSMAITRIYADHNMLGSPKERAISMLEGAVLDIRYVSRIATSKDEETLKV 120
DB 61 GDVLTOSMAITRIYADHNMLGSPKERAISMLEGAVLDIRYVSRIATSKDEETLKV 120
QY 121 DFLSKLPMLKMFEDRLSHKTYTLNGDHYTHDPFMLYDALDVLYNDPMLCLDAFPPLVSEFK 180
DB 121 DFLSKLPMLKMFEDRLSHKTYTLNGDHYTHDPFMLYDALDVLYNDPMLCLDAFPPLVSEFK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFFGGGDHPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFFGGGDHPK 218

RESULT 7
US-09-844-988-6
Sequence 6, Application US/09844988
Patent No. US20020158764A1
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
Zhu, Hengyi
Barbosa, Miguel
Li, Gfan
Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,988
FILING DATE: 26-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/910,820
FILING DATE: 1997-08-13
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-844-988-6

Query Match 98.7% Score 1149; DB 9; Length 272;
Best Local Similarity 98.6%; Pred. No. 1.3e-104;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGKIRKGLVQPTLLLEYLEEKYEHLHYERDEGDKMKNKFFELGLEFPNLPYYID 60
DB 1 MSPILGKIRKGLVQPTLLLEYLEEKYEHLHYERDEGDKMKNKFFELGLEFPNLPYYID 60
QY 61 GDVKLTQSMATIRYADKHNMLGSGPKRAEISMLEGAVLDIRGVSRIVASKDEFTLKV 120
DB 61 GDVKLTQSMATIRYADKHNMLGSGPKRAEISMLEGAVLDIRGVSRIVASKDEFTLKV 120
QY 121 DFLSKLPMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPMLCLDAFPRLVSFK 180
DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLYMDPMLCLDAFPRLVSFK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFGGDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFGGDHPPK 218

RESULT 8
US-09-844-908-4
Sequence 4, Application US/09844908
Patent No. US20020151021A1
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
Zhu, Hengyi
Barbosa, Miguel
Li, Gian
Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,908
FILING DATE: 27-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-844-908-4

Query Match 98.7% Score 1149; DB 10; Length 272;
Best Local Similarity 98.6%; Pred. No. 1.3e-104;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGKIRKGLVQPTLLLEYLEEKYEHLHYERDEGDKMKNKFFELGLEFPNLPYYID 60
DB 1 MSPILGKIRKGLVQPTLLLEYLEEKYEHLHYERDEGDKMKNKFFELGLEFPNLPYYID 60
QY 61 GDVKLTQSMATIRYADKHNMLGSGPKRAEISMLEGAVLDIRGVSRIVASKDEFTLKV 120
DB 61 GDVKLTQSMATIRYADKHNMLGSGPKRAEISMLEGAVLDIRGVSRIVASKDEFTLKV 120
QY 121 DFLSKLPMLKMFEDRLSHKTYLNGDHTHPDFMLYDALDVLYMDPMLCLDAFPRLVSFK 180
DB 121 DFLSKLPMLKMFEDRLCHKTYLNGDHTHPDFMLYDALDVLYMDPMLCLDAFPRLVSFK 180
QY 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFGGDHPPK 218
DB 181 KRIEAIPOIDKYLKSSKIYAMPLOGMQATFGGDHPPK 218

RESULT 9
US-09-844-908-6
Sequence 6, Application US/09844908
Patent No. US20020151021A1
GENERAL INFORMATION:
APPLICANT: Mercurio, Frank
Zhu, Hengyi
Barbosa, Miguel
Li, Gian
Murray, Brian W.
TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,908
FILING DATE: 27-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 272 amino acids
TYPE: amino acid

Tue May 20 13:19:23 2003

us-10-081-408-5.rapb

Page

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-09-844-988-6

Query Match 98.7%; Score 1149; DB 10; Length 272;

Best Local Similarity 98.6%; Pred. No. 1.3e-104;

Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
OY 1 MSPILGWKIKGLVPTLLLEYLEEKYEHEHLYERDEGDKWKKFELGLEFPNLPYYID 60
DB 1 MSPILGWKIKGLVPTLLLEYLEEKYEHEHLYERDEGDKWKKFELGLEFPNLPYYID 60
OY 61 GVKLTQSAIIRYIADKHNMLGSPKERAISMLEGAVLDIRYGSRIAYSKEDETLKV 120
DB 61 GVKLTQSAIIRYIADKHNMLGSPKERAISMLEGAVLDIRYGSRIAYSKEDETLKV 120
OY 121 DFLSKLPEMLKMFEDRLSKHTYLNQDHYTHPDFMLYDALDVLYMDPKCLDAFPKLVSPK 180
DB 121 DFLSKLPEMLKMFEDRLSKHTYLNQDHYTHPDFMLYDALDVLYMDPKCLDAFPKLVSPK 180
OY 181 KRLEAIPOIDKYLKSSKYIAMPLOGQWQATFGGDDHPK 218
DB 181 KRLEAIPOIDKYLKSSKYIAMPLOGQWQATFGGDDHPK 218
```

RESULT 10

US-09-844-988-3

Sequence 3, Application US/09844988

Patent No. US20020158764A1

GENERAL INFORMATION:

APPLICANT: Mercurio, Frank

Zhu, Hengyi

Barbosa, Miguel

Li, Gian

Murray, Brian W.

TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE

COMPLEX AND METHODS OF USE THEREFOR

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESSES:

ADDRESS: SEED and BERRY LLP

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/844,988

FILING DATE: 26-Apr-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/910,820

FILING DATE: 1997-08-13

ATTORNEY/AGENT INFORMATION:

NAME: Makl, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 860098.413C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 282 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 3:

US-09-844-988-3

Query Match 98.7%; Score 1149; DB 9; Length 282;

Best Local Similarity 98.6%; Pred. No. 1.4e-104;

Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
OY 1 MSPILGWKIKGLVPTLLLEYLEEKYEHEHLYERDEGDKWKKFELGLEFPNLPYYID 60
DB 1 MSPILGWKIKGLVPTLLLEYLEEKYEHEHLYERDEGDKWKKFELGLEFPNLPYYID 60
OY 61 GVKLTQSAIIRYIADKHNMLGSPKERAISMLEGAVLDIRYGSRIAYSKEDETLKV 120
DB 61 GVKLTQSAIIRYIADKHNMLGSPKERAISMLEGAVLDIRYGSRIAYSKEDETLKV 120
OY 121 DFLSKLPEMLKMFEDRLSKHTYLNQDHYTHPDFMLYDALDVLYMDPKCLDAFPKLVSPK 180
DB 121 DFLSKLPEMLKMFEDRLSKHTYLNQDHYTHPDFMLYDALDVLYMDPKCLDAFPKLVSPK 180
OY 181 KRLEAIPOIDKYLKSSKYIAMPLOGQWQATFGGDDHPK 218
DB 181 KRLEAIPOIDKYLKSSKYIAMPLOGQWQATFGGDDHPK 218
```

RESULT 11

US-09-844-988-5

Sequence 5, Application US/09844988

Patent No. US20020158764A1

GENERAL INFORMATION:

APPLICANT: Mercurio, Frank

Zhu, Hengyi

Barbosa, Miguel

Li, Gian

Murray, Brian W.

TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE

COMPLEX AND METHODS OF USE THEREFOR

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESSES:

ADDRESS: SEED and BERRY LLP

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/844,988

FILING DATE: 26-Apr-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/910,820

FILING DATE: 1997-08-13

ATTORNEY/AGENT INFORMATION:

NAME: Makl, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 860098.413C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 282 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-09-844-988-5

Query Match 98.7%; Score 1149; DB 9; Length 282;

Best Local Similarity 98.6%; Pred. No. 1.4e-104;

Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGWKIKGLVPTLLEYLEEKEEHLERDEGDKMKNKFFELGEPNLPYYID 60
DB 1 MSPILGWKIKGLVPTLLEYLEEKEEHLERDEGDKMKNKFFELGEPNLPYYID 60
QY 61 GVKLTQSMALIRYIADKHNMLGCGPKERAETSMLEGAVALDIRYGVSRVIAYSKDEFTLVK 120
DB 61 GVKLTQSMALIRYIADKHNMLGCGPKERAETSMLEGAVALDIRYGVSRVIAYSKDEFTLVK 120
QY 121 DFLSLPELKKFEEDRLSHKTYLNGDHYTHPDPMLYDALDVLYLMDPKCLDAFPPLVYSFK 180
DB 121 DFLSLPELKKFEEDRLSHKTYLNGDHYTHPDPMLYDALDVLYLMDPKCLDAFPPLVYSFK 180
QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMQATFFGGGDHPK 218
DB 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMQATFFGGGDHPK 218

RESULT 12

US-09-844-908-3
Sequence 3, Application US/09844908
Patent No. US20020151021A1
GENERAL INFORMATION:

APPLICANT: Mercurio, Frank

Zhu, Hengyi
Barbosa, Miguel
Li, Gfan

Murray, Brian W.

TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:
ADDRESS: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA

ZIP: 98104

COUNTRY: USA

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,908
FILING DATE: 27-Apr-2001
CLASSIFICATION: <Unknown>

Prior Application Data:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997

ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1

TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-844-908-3

Query Match 98.7%, Score 1149; DB 10; Length 282;
Best Local Similarity 98.6%; Pred. No. 1.4e-104;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGWKIKGLVPTLLEYLEEKEEHLERDEGDKMKNKFFELGEPNLPYYID 60
DB 1 MSPILGWKIKGLVPTLLEYLEEKEEHLERDEGDKMKNKFFELGEPNLPYYID 60

QY 61 GVKLTQSMALIRYIADKHNMLGCGPKERAETSMLEGAVALDIRYGVSRVIAYSKDEFTLVK 120
DB 61 GVKLTQSMALIRYIADKHNMLGCGPKERAETSMLEGAVALDIRYGVSRVIAYSKDEFTLVK 120
QY 121 DFLSLPELKKFEEDRLSHKTYLNGDHYTHPDPMLYDALDVLYLMDPKCLDAFPPLVYSFK 180
DB 121 DFLSLPELKKFEEDRLSHKTYLNGDHYTHPDPMLYDALDVLYLMDPKCLDAFPPLVYSFK 180
QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMQATFFGGGDHPK 218
DB 181 KRIEAIPOIDKYLKSSKXYIAMPLOGMQATFFGGGDHPK 218

RESULT 13

US-09-844-908-5
Sequence 5, Application US/09844908
Patent No. US20020151021A1
GENERAL INFORMATION:

APPLICANT: Mercurio, Frank

Zhu, Hengyi
Barbosa, Miguel
Li, Gfan

Murray, Brian W.

TITLE OF INVENTION: STIMULUS-INDUCIBLE PROTEIN KINASE
COMPLEX AND METHODS OF USE THEREFOR

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:
ADDRESS: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA

ZIP: 98104

COUNTRY: USA

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/844,908
FILING DATE: 27-Apr-2001
CLASSIFICATION: <Unknown>

Prior Application Data:
APPLICATION NUMBER: US/08/910,820
FILING DATE: 12-AUG-1997

ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 860098.413C1

TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-844-908-5

Query Match 98.7%, Score 1149; DB 10; Length 282;
Best Local Similarity 98.6%; Pred. No. 1.4e-104;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGWKIKGLVPTLLEYLEEKEEHLERDEGDKMKNKFFELGEPNLPYYID 60
DB 1 MSPILGWKIKGLVPTLLEYLEEKEEHLERDEGDKMKNKFFELGEPNLPYYID 60
QY 61 GVKLTQSMALIRYIADKHNMLGCGPKERAETSMLEGAVALDIRYGVSRVIAYSKDEFTLVK 180
DB 61 GVKLTQSMALIRYIADKHNMLGCGPKERAETSMLEGAVALDIRYGVSRVIAYSKDEFTLVK 180

QY 121 DFLSKLPMLKMFEDRLSHKTYTLNGDHYTHDPFMDLYDALDVLVLYMDPCLDAFPKLVSEK 180
 |||||
 Db 121 DFLSKLPMLKMFEDRLSHKTYTLNGDHYTHDPFMDLYDALDVLVLYMDPCLDAFPKLVSEK 180
 |||||
 QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGDHPPK 218
 |||||
 Db 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGDHPPK 218

RESULT 14

US-09-910-600-26
 ; Sequence 26, Application US/09910600
 ; Publication No. US2003003631A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Longphre, Malinda
 ; APPLICANT: Chang, Han
 ; APPLICANT: Whitley, Gena
 ; TITLE OF INVENTION: NOVEL SINGLETS AND USES THEREOF
 ; FILE REFERENCE: D0003NP
 ; CURRENT APPLICATION NUMBER: US/09/910,600
 ; CURRENT FILING DATE: 2001-07-20
 ; PRIOR APPLICATION NUMBER: 60/220,139
 ; PRIOR FILING DATE: 2000-07-21
 ; NUMBER OF SEQ ID NOS: 32
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 26
 ; LENGTH: 298
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: L3cyto-Y641
 ;
 US-09-910-600-26

Query Match 98.7%; Score 1149; DB 9; Length 298;
 Best Local Similarity 98.6%; Pred. No. 1.5e-104;

Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGWKIKGLVOPTRLLLEYLEEKYEHLRYERDEGDKWRNKKFELGLEFPNLPYYID 60
 |||||
 Db 1 MSPILGWKIKGLVOPTRLLLEYLEEKYEHLRYERDEGDKWRNKKFELGLEFPNLPYYID 60
 |||||
 QY 61 GDVKLTOSMAIIRYIADKHNMLGGSPKRAEISMLEGAVLDIRYGVSRAYSKDPEITLV 120
 |||||
 Db 61 GDVKLTOSMAIIRYIADKHNMLGGSPKRAEISMLEGAVLDIRYGVSRAYSKDPEITLV 120
 |||||
 QY 121 DFLSKLPMLKMFEDRLSHKTYTLNGDHYTHDPFMDLYDALDVLVLYMDPCLDAFPKLVSEK 180
 |||||
 Db 121 DFLSKLPMLKMFEDRLSHKTYTLNGDHYTHDPFMDLYDALDVLVLYMDPCLDAFPKLVSEK 180
 |||||
 QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGDHPPK 218
 |||||
 Db 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGDHPPK 218

RESULT 15

US-10-267-311-25
 ; Sequence 25, Application US/10267311
 ; Publication No. US20030050469A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Siegel, Marvin
 ; APPLICANT: Chu, N. Randall
 ; APPLICANT: Mlzen, Lee A.
 ; TITLE OF INVENTION: INDUCTION OF A TH1-LIKE RESPONSE IN VITRO
 ; FILE REFERENCE: 12071/002001
 ; CURRENT APPLICATION NUMBER: US/10/267,311
 ; CURRENT FILING DATE: 2002-10-09
 ; PRIOR APPLICATION NUMBER: US/09/613,303
 ; PRIOR FILING DATE: 2000-07-10
 ; PRIOR APPLICATION NUMBER: US 60/143,757
 ; PRIOR FILING DATE: 1999-07-08
 ; NUMBER OF SEQ ID NOS: 55

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 25
 ; LENGTH: 324
 ; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:
 ; OTHER INFORMATION: fusion sequence

US-10-267-311-25

Query Match 98.7%; Score 1149; DB 9; Length 324;
 Best Local Similarity 98.6%; Pred. No. 1.7e-104;

Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGWKIKGLVOPTRLLLEYLEEKYEHLRYERDEGDKWRNKKFELGLEFPNLPYYID 60
 |||||
 Db 1 MSPILGWKIKGLVOPTRLLLEYLEEKYEHLRYERDEGDKWRNKKFELGLEFPNLPYYID 60
 |||||
 QY 61 GDVKLTOSMAIIRYIADKHNMLGGSPKRAEISMLEGAVLDIRYGVSRAYSKDPEITLV 120
 |||||
 Db 61 GDVKLTOSMAIIRYIADKHNMLGGSPKRAEISMLEGAVLDIRYGVSRAYSKDPEITLV 120
 |||||
 QY 121 DFLSKLPMLKMFEDRLSHKTYTLNGDHYTHDPFMDLYDALDVLVLYMDPCLDAFPKLVSEK 180
 |||||
 Db 121 DFLSKLPMLKMFEDRLSHKTYTLNGDHYTHDPFMDLYDALDVLVLYMDPCLDAFPKLVSEK 180
 |||||
 QY 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGDHPPK 218
 |||||
 Db 181 KRIEAIPOIDKYLKSSKXYIAMPLOGWQATFGGDHPPK 218

Search completed: May 20, 2003, 12:02:22

Job time : 35.7335 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 08:02:28 ; Search time 5.14571 Seconds

(Without alignments)
1757.161 Million cell updates/sec

Title: US-10-081-408-5

Perfect score: 1164

Sequence: 1 MSPILGYWKIKGLVQPTRL.....IAMPLOGQWATGCGDHPK 218

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 segs, 41476328 residues

112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1149	98.7	218	1	GT26_SCHJA
2	990	85.1	218	1	GT27_SCHJA
3	927	79.6	218	1	GT26_SCHJA
4	717.5	61.6	220	1	GT29_FASHE
5	708	60.8	217	1	GT27_FASHE
6	690	59.3	217	1	GT28_FASHE
7	689	59.2	217	1	GT26_FASHE
8	489.5	42.1	217	1	GTM1_MOUSE
9	489.5	42.1	217	1	GTM1_MOUSE
10	488.5	42.0	218	1	GTM1_MOUSE
11	486.5	41.8	218	1	GTM6_MOUSE
12	485.5	41.7	217	1	GTM6_MOUSE
13	484.5	41.6	217	1	GTM2_HUMAN
14	481.5	41.4	217	1	GTM2_HUMAN
15	476.5	40.9	217	1	GTM2_HUMAN
16	476.5	40.9	217	1	GTM2_HUMAN
17	475.5	40.9	217	1	GTM2_HUMAN
18	473.5	40.7	217	1	GTM2_HUMAN
19	471.5	40.5	217	1	GTM2_HUMAN
20	469.5	40.3	217	1	GTM2_HUMAN
21	463.5	39.8	217	1	GTM2_HUMAN
22	460.5	39.6	217	1	GTM2_HUMAN
23	454.5	39.0	224	1	GTM3_HUMAN
24	443.5	38.1	224	1	GTM3_HUMAN
25	440.5	37.8	219	1	GTM2_CHICK
26	394.5	33.9	219	1	GTM2_CHICK
27	359.5	22.3	208	1	GTPL_MOUSE
28	254.5	21.9	208	1	GTPL_MOUSE
29	253	21.9	208	1	GTPL_MOUSE
30	252	21.6	207	1	GTPL_MOUSE
31	251	21.6	207	1	GTPL_MOUSE
32	249	21.4	209	1	GTPL_MOUSE
33	247	21.2	209	1	GTPL_MOUSE

34	247	21.2	209	1	GTPL_MOUSE	P47954 cricetus
35	246	21.1	209	1	GTPL_MOUSE	P04906 ratius norv
36	245	21.0	209	1	GTPL_MOUSE	P46424 cricetus
37	244	21.0	209	1	GTPL_MOUSE	P28801 bos taurus
38	243	20.9	209	1	GTPL_MOUSE	P28801 bos taurus
39	243	20.5	209	1	GTPL_MOUSE	P28801 bos taurus
40	226.5	19.5	208	1	GTPL_MOUSE	P10289 caenorhabd1
41	221.5	19.0	223	1	GTAL_MOUSE	P08863 oryctolagus
42	213.5	18.3	210	1	GTAL_MOUSE	P83325 bufo bufo
43	209.5	18.0	221	1	GTAL_MOUSE	P04903 ratius norv
44	207.5	17.8	221	1	GTAL_MOUSE	P10648 mus musculu
45	207.5	17.8	222	1	GTAL_MOUSE	P13745 mus musculu

ALIGNMENTS

RESULT 1	GT26_SCHJA	STANDARD;	PRT;	218 AA.
ID	GT26_SCHJA	STANDARD;	PRT;	218 AA.
AC	P08515;			
DT	01-AUG-1988 (Rel. 08, Created)			
DT	01-AUG-1988 (Rel. 08, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	Glutathione S-transferase 26 kDa (EC 2.5.1.18) (GST 26) (S26 antigen)			
DE	(GST class-alpha).			
OS	Schistosoma japonicum (Blood fluke).			
OC	Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;			
OC	Schistosomatidae; Schistosomatidae; Schistosoma.			
OX	NCBI_Taxid=6182;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RF	MEDLINE=87041520; PubMed=3095841;			
RA	Smith D.B., Davern K.M., Board P.G., Tiu W.U., Garcia E.G.,			
RA	Mitchell G.F.;			
RA	Proc. Natl. Acad. Sci. U.S.A. 84:6541-6541(1987).			
RP	[3]			
RN	X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).			
RF	MEDLINE=95276631; PubMed=7538846;			
RA	Lim K., Ho J.X., Keeling K., Gillingland G.L., Ji X., Ruker F.,			
RA	Carter D.C.;			
RT	"Three-dimensional structure of Schistosoma japonicum glutathione S-			
RT	transferase fused with a six-amino acid conserved neutralizing			
RT	epitope of gp41 from HIV."			
RL	Protein Sci. 3:2233-2244(1994).			
RP	[4]			
RF	X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).			
RF	MEDLINE=95156484; PubMed=7853399;			
RA	McGlynn M.A., Williams D.R., Tainer J.A.;			
RT	"Crystal structures of a schistosomal drug and vaccine target:			
RT	glutathione S-transferase from Schistosoma japonica and its complex			
RT	with the leading antischistosomal drug praziquantel."			
RL	J. Mol. Biol. 246:21-27(1995).			
CC	-1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER			
CC	OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.			
CC	-1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE			
CC	PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO			
CC	SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF			
CC	HAEMATIN IN THE PARASITE GUT.			
CC	-1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.			
CC	-1- SUBUNIT: HOMODIMER.			
CC	-1- MISCELLANEOUS: THERE ARE AT LEAST TWO ISOENZYMES OF GST IN			
CC	S. JAPONICUM.			
CC	-1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M14654; AAB59203.1; -
DR PIR; A26484; A26484.
DR PDB; 1GNE; 30-NOV-94.
DR PDB; 1GTR; 07-FEB-95.
DR PDB; 1GTR; 01-DEC-95.
DR PDB; 1B8X; 12-APR-99.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR004045; GST_Nterm.
DR Pfam; PF00043; GST_C; 1.
DR Pfam; PF02798; GST_N; 1.
DR Transferrase; Antigen; Multigene family; 3D-structure.
KW SEQUENCE 218 AA; 25498 MW; 5E2AC418BDE0F13F CRC64;

Query Match 98.7%; Score 1149; DB 1; Length 218;
Best Local Similarity 98.6%; Pred. No. 3.8e-90;
Matches 215; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVQPRLLLEYLEEKEEYERDEGDKWRNKKFELGLEFPNLPYYID 60
DB 1 MSPILGYWKIKGLVQPRLLLEYLEEKEEYERDEGDKWRNKKFELGLEFPNLPYYID 60
QY 61 GDVLTQSMALIRYIADKHNLMGSPKERAISMLEGAVDIRYGSRIAYSKDFETLKY 120
DB 61 GDVLTQSMALIRYIADKHNLMGSPKERAISMLEGAVDIRYGSRIAYSKDFETLKY 120
QY 121 DFLSKLEPMKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVLVYMDPCDAPFKLYSFK 180
DB 121 DFLSKLEPMKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVLVYMDPCDAPFKLYSFK 180
QY 121 DFLSKLEPMKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVLVYMDPCDAPFKLYSFK 180
DB 121 DFLSKLEPMKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVLVYMDPCDAPFKLYSFK 180
QY 181 KRIPAIPOIDKYLKSKSYIAMPLOGMOATFGGDPHPK 218
DB 181 KRIPAIPOIDKYLKSKSYIAMPLOGMOATFGGDPHPK 218

RESULT 2
GT27_SCHMA STANDARD; PRT; 218 AA.
ID GT27_SCHMA STANDARD; PRT; 218 AA.
AC P35661;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Glutathione S-transferase 26 kDa (EC 2.5.1.18) (GST 26) (SM26/2)
DE antigen) (GST class-alpha).
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6183;

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92131046; Pubmed-1775156;
RA Wright M.D., Harrison R.A., Melder A.M., Newport G.R., Mitchell G.F.;
RT "Another 26-kilodalton glutathione S-transferase of Schistosoma
mansoni".
RL Mol. Biochem. Parasitol. 49:177-179(1991).
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
CC HAEMATIN IN THE PARASITE GUT.
CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M73624; -; NOT_ANNOTATED_CDS.
DR PIR; A45556; A45556.
DR HSSP; P08515; 1GTA.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR004045; GST_Nterm.
DR Pfam; PF00043; GST_C; 1.
DR Pfam; PF02798; GST_N; 1.
DR Transferrase; Antigen; Multigene family.
KW SEQUENCE 218 AA; 25411 MW; D8D3EE9028B36185 CRC64;

Query Match 85.1%; Score 990; DB 1; Length 218;
Best Local Similarity 83.0%; Pred. No. 1e-76;
Matches 181; Conservative 19; Mismatches 18; Indels 0; Gaps 0;

QY 1 MSPILGYWKIKGLVQPRLLLEYLEEKEEYERDEGDKWRNKKFELGLEFPNLPYYID 60
DB 1 MAPKLGTYWKIKGLVQPRLLLEYLEEKEEYERDEGDKWRNKKFELGLEFPNLPYYID 60
QY 61 GDVLTQSMALIRYIADKHNLMGSPKERAISMLEGAVDIRYGSRIAYSKDFETLKY 120
DB 61 GDVLTQSMALIRYIADKHNLMGSPKERAISMLEGAVDIRYGSRIAYSKDFETLKY 120
QY 121 DFLSKLEPMKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVLVYMDPCDAPFKLYSFK 180
DB 121 DFLSKLEPMKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVLVYMDPCDAPFKLYSFK 180
QY 121 DFLSKLEPMKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVLVYMDPCDAPFKLYSFK 180
DB 121 DFLSKLEPMKMFEDRLSHKTYLNGDHYTHPDFMLYDALDVLVYMDPCDAPFKLYSFK 180
QY 181 KRIPAIPOIDKYLKSKSYIAMPLOGMOATFGGDPHPK 218
DB 181 KRIPAIPOIDKYLKSKSYIAMPLOGMOATFGGDPHPK 218

RESULT 3
GT26_SCHMA STANDARD; PRT; 218 AA.
ID GT26_SCHMA STANDARD; PRT; 218 AA.
AC P15964;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Glutathione S-transferase 26 kDa (EC 2.5.1.18) (GST 26) (SM26/1)
DE antigen) (GST class-alpha).
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6183;

RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-Puerto Rican;
RX MEDLINE-90348716; Pubmed-2385266;
RA Trotein F., Kieny M.P., Vervaeke C., Torgler G., Pierce R.J.,
RA Balloul J.-M., Schmitt D., Lecocq J.-P., Capron A.;
RT "Molecular cloning and tissue distribution of a 26-kilodalton
Schistosoma mansoni glutathione S-transferase".
RL Mol. Biochem. Parasitol. 41:35-44(1990).
CC [2]
CC SEQUENCE OF 8-218 FROM N.A.
CC STRAIN-Puerto Rican;
CC MEDLINE-90271935; Pubmed-1693415;
CC Henkle K.J., Davern K.M., Wright M.D., Ramos A.J., Mitchell G.F.;
CC "Comparison of the cloned genes of the 26- and 28-kilodalton
CC glutathione S-transferases of Schistosoma japonicum and Schistosoma
CC mansoni".
CC RL Mol. Biochem. Parasitol. 40:23-34(1990).
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF


```

CC HAEMATIN IN THE PARASITE GUT.
CC -1- CATALYTIC ACTIVITY: RX + glutathione - HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- TISSUE SPECIFICITY: TEGUMENT AND IN SUBSEGMENTARY PARENCHYMAL CELLS. GST 26 MAY BE ACTIVELY EXCRETED BY ADULT WORMS.
CC -1- MISCELLANEOUS: THERE ARE AT LEAST THREE ISOENZYMES OF GST IN S. MANSONI.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M31106; AAA28888.1; -.
DR EMBL: M26913; AAA28889.1; -.
DR PIR: A45523; A45523.
DR HSSP: P08515; 1GTA.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C; 1.
DR Pfam: PF02798; GST_N; 1.
DR Transferrase; Antigen; Multigene family.
KW SEQUENCE 218 AA; 25401 MW; 061A6548A842D6E8 CRC64;
SQ
Query Match 79.6%; Score 927; DB 1; Length 218;
Best Local Similarity 80.3%; Pred. No. 2,1e-71;
Matches 175; Conservative 15; Mismatches 28; Indels 0; Gaps 0;
QY 1 MSPILGYWKIKGLVQPTLLLEDEKYEHLERDEGDKRNKKEFELGEPNLPYYID 60
DB 1 MAPKFGYWKVKGVLQPTRLLEHLEETEYERAYDRNEIDAMSNKFKLGEPNLPYYID 60
QY 61 GDVAVLTQSMALIRIADKHNHMGSPKRAEISMLEGAVLDIRGVSNIAVSKDEFTKY 120
DB 61 GDFLKTOSMAIRIYADKHNHMGACPKERAEISMLEGAVLDIRGVSNIAVSKDEFTKY 120
QY 121 DFLSKPLMKAFEDRLSKHTYTLNDGHTHPDPMYDALDVLYMDPNCLOAFPLVSFK 180
DB 121 DFLNKLPEGLKMFEDRLSKHTYTLNDGHTHPDPMYDALDVLYMDSCLEFPPLVSFK 180
QY 181 KRIPAIPOIDRYLKSSKRYIAMPLOGMOATFGGCDHPK 218
DB 181 KCIEDLPQIKNYLNNSRYIKWPLOGMDATFGGCDTPPK 218

RESULT 4
GT29_FASHE STANDARD; PRT; 220 AA.
AC P56598;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutathione S-transferase 26 kDa 1 (EC 2.5.1.18) (GST1) (FHL) (GST class-alpha).
OS Fasciola hepatica (liver fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Fasciolidae; Fasciola.
OC Echinostomida; Echinostomata; Fasciolidae; Fasciolidae; Fasciola.
OX NCBI_TaxID=6192;
RN
RP SEQUENCE FROM N.A.
RA MEDLINE=92155306; PubMed=1740183;
RT Paracicio M., Wilson L.R., Cramer S.L., Wijffels G.L., Spithill T.W.; "Molecular characterization of cDNA sequences encoding glutathione S-transferases of Fasciola hepatica."
RT Exp. Parasitol. 74:232-237(1992).
RN [2]
RP ERRATUM.
RX MEDLINE=94039664; PubMed=8224094;
RA Paracicio M., Wilson L.R., Cramer S.L., Wijffels G.L., Spithill T.W.;

```

```

RL Exp. Parasitol. 77:385-385(1993).
RN [3]
RA SEQUENCE OF 22-220 FROM N.A.
RP Cramer S.;
RL Patient number W09008819, 09-AUG-1990.
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF HAEMATIN IN THE PARASITE GUT.
CC -1- CATALYTIC ACTIVITY: RX + glutathione - HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: A00993; CA00118.1; -.
DR HSSP: P31670; 1PHE.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C; 1.
DR Pfam: PF02798; GST_N; 1.
DR Transferrase; Antigen; Multigene family.
KW INT_MET 0 0 BY SIMILARITY.
FT CONFLICT 22 22 Y -> V (IN REF. 3).
FT CONFLICT 110 111 DP -> VS (IN REF. 3).
FT CONFLICT 189 189 A -> P (IN REF. 3).
SQ SEQUENCE 220 AA; 25598 MW; 27B9F150B750101F CRC64;

Query Match 61.6%; Score 717.5; DB 1; Length 220;
Best Local Similarity 60.6%; Pred. No. 1e-53;
Matches 129; Conservative 32; Mismatches 51; Indels 1; Gaps 1;
QY 5 LGYWKIKGLVQPTLLLEDEKYEHLERDEGDKRNKKEFELGEPNLPYYIDGVK 64
DB 4 LGYWKIKGLVQPTLLLEDEKYEHLERDEGDKRNKKEFELGEPNLPYYIDGVK 62
QY 65 LTQSMALIRIADKHNHMGSPKRAEISMLEGAVLDIRGVSNIAVSKDEFTKYVFLS 124
DB 63 LTQSMALIRIADKHNHMGSPKRAEISMLEGAVLDIRGVSNIAVSKDEFTKYVFLS 122
QY 125 KLPEKLMKFEEDRLSKHTYTLNDGHTHPDPMYDALDVLYMDPNCLOAFPLVSFKRIE 184
DB 123 KLPEKLMKFEEDRLSKHTYTLNDGHTHPDPMYDALDVLYMDPNCLOAFPLVSFKRIE 182
QY 185 AIPQIDRYLKSSKRYIAMPLOGMOATFGGCDHPK 217
DB 183 ALPSIKAYMESNRIFKWPLOGMDATFGGCDAPK 215

RESULT 5
GT27_FASHE STANDARD; PRT; 217 AA.
ID GT27_FASHE
AC P31670;
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutathione S-transferase 26 kDa 47 (EC 2.5.1.18) (GST47) (FH47)
DE (GST class-alpha).
OS Fasciola hepatica (liver fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Fasciolidae; Fasciola.
OC Echinostomida; Echinostomata; Fasciolidae; Fasciolidae; Fasciola.
OX NCBI_TaxID=6192;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92155306; PubMed=1740183;

```

RA Panaccio M., Wilson L.R., Cramerl S.L., Wijffels G.L., Spithill T.W.;
 RT "Molecular characterization of cDNA sequences encoding glutathione S-
 RT transferases of Fasciola hepatica.";
 RL Exp. Parasitol. 74:232-237(1992).
 RN [2]
 RP ERRATUM.
 RX MEDLINE=94039664; PubMed=8224094;
 RA Panaccio M., Wilson L.R., Cramerl S.L., Wijffels G.L., Spithill T.W.;
 RL Exp. Parasitol. 77:385-385(1993).
 RN [3]
 RP SEQUENCE OF 7-105 FROM N.A.
 RA Cramerl S.;
 RL Patent number W09008819, 09-AUG-1990.
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).
 RX MEDLINE=98035725; PubMed=9367777;
 RA Rosjohn J., Fell S.C., Milce M.C.J., Sexton J.L., Spithill T.W.,
 RA Parker M.W.;
 RT "Crystallization, structural determination and analysis of a novel
 RT parasite vaccine candidate: Fasciola hepatica glutathione
 RT S-transferase.";
 RL J. Mol. Biol. 273:857-872(1997).
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
 CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
 CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
 CC HAEMATIN IN THE PARASITE GUT.
 CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; M77681; AAA29140.1; -;
 DR EMBL; A00996; CA00121.1; -;
 DR PDB; 1FHE; 29-JUL-98.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam; PF00043; GST_C; 1.
 DR Pfam; PF02798; GST_N; 1.
 DR Transferrase; Antigen: Multigene family; 3D-structure.
 KW INT_MET 0
 FT CONFLICT 65 65 T -> I (IN REF. 3).
 FT CONFLICT 102 105 RIGF -> FEEL (IN REF. 3).
 SQ SEQUENCE 217 AA; 25281 MW; 0FB8BFE63029E03 CRC64;
 Query Match 60.8%; Score 708; DB 1; Length 217;
 Best Local Similarity 58.2%; Pred. No. 6.5e-53;
 Matches 124; Conservative 37; Mismatches 52; Indels 0; Gaps 0;
 Oy 5 LGYWKIKGLVPTRLLEYLEEKYEHLXERDEGKMRKKKELEFPNLYYIDGVK 64
 Db 4 LGYWKIKGLVPTRLLEYLEEKYEHLXERDEGKMRKKKELEFPNLYYIDGVK 63
 Oy 65 LTQSAIIRIYADKHNMLGSGPKERAEISMLEGAVLDIRYGVSRAYSKEFTLVDFLS 124
 Db 64 LTQSAIIRIYADKHNMLGSGTPREARARISMIGAMDLRIGGRGYCNKFEVVAEEYK 123
 Oy 125 KLPEMLKFEEDRLSHKTYLNGDVTHTPDMYLYDALDYYLYMDPKCLDAFPKLVSEKRIE 184
 Db 124 ELPKTLKMSDFLDGRHLYLTGSSVSHVDFMLYETLDSIRYLAIPHCLDEPRKXKEKRSRIE 183
 Oy 185 AIPQIDKXLSKSYTAMPLOGQWATFFGGDHP 217
 Db 184 ALPKIKAYMESKRFITKWLPLNGMAASFAGADAP 216

RESULT 6
 GT28_FASHE STANDARD; PRT; 217 AA.
 AC P31671;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Glutathione S-transferase 26 kDa 7 (EC 2.5.1.18) (GST7) (FH7) (GST
 DE class-alpha).
 OS Fasciola hepatica (Liver fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 OC Echinostomida; Echinostomata; Fasciolidae; Fasciola.
 OX NCBI_Taxid=6192;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92155306; PubMed=1740183;
 RA Panaccio M., Wilson L.R., Cramerl S.L., Wijffels G.L., Spithill T.W.;
 RT "Molecular characterization of cDNA sequences encoding glutathione S-
 RT transferases of Fasciola hepatica.";
 RL Exp. Parasitol. 74:232-237(1992).
 RN [2]
 RP ERRATUM.
 RX MEDLINE=94039664; PubMed=8224094;
 RA Panaccio M., Wilson L.R., Cramerl S.L., Wijffels G.L., Spithill T.W.;
 RL Exp. Parasitol. 77:385-385(1993).
 RN [3]
 RP SEQUENCE OF 8-217 FROM N.A.
 RA Cramerl S.;
 RL Patent number W09008819, 09-AUG-1990.
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
 CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
 CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
 CC HAEMATIN IN THE PARASITE GUT.
 CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; M77680; AAA29139.1; -;
 DR EMBL; A00994; CA00119.1; -;
 DR HSRP; P31670; 1FHE.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam; PF00043; GST_C; 1.
 DR Pfam; PF02798; GST_N; 1.
 DR Transferrase; Antigen: Multigene family.
 KW INT_MET 0
 FT CONFLICT 35 42 NDREKWLQ -> MGRNGMA (IN REF. 3).
 FT CONFLICT 188 196 IKYWKSR -> SYMSRA (IN REF. 3).
 SQ SEQUENCE 217 AA; 25196 MW; 0099ELF59E49A49E CRC64;
 Query Match 59.3%; Score 690; DB 1; Length 217;
 Best Local Similarity 58.7%; Pred. No. 2.1e-51;
 Matches 125; Conservative 30; Mismatches 58; Indels 0; Gaps 0;
 Oy 5 LGYWKIKGLVPTRLLEYLEEKYEHLXERDEGKMRKKKELEFPNLYYIDGVK 64
 Db 4 LGYWKIKGLVPTRLLEYLEEKYEHLXERDEGKMRKKKELEFPNLYYIDGVK 63
 Oy 65 LTQSAIIRIYADKHNMLGSGPKERAEISMLEGAVLDIRYGVSRAYSKEFTLVDFLS 124
 Db 64 LTQSAIIRIYADKHNMLGSGTPREARARISMIGAMDLRIGGLTCYKPKFEELKGYLK 123

```

OY 125 KLPEELKNEEDRLSKRTKLDNDGHTVHPFMYLADVDLVDPKDLDPKLVSEFKRIE 184
Db 124 GLPTLLKMSDFDLGRDROLICSSVSHVDPMYREALDCLRILAPOCLDLPFKLKEKRNIE 183
OY 185 AIPQIDKYLKSSKYIAMPLOGMOWATFGGGDHP 217
Db 184 DLPKIKEXMKSERFIKMPDLHSWTSFPGCGDAP 216

RESULT 7
GT26_FASHE STANDARD; PRT; 217 AA.
AC P30112;
DT 01-APR-1993 (Rel. 25, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutathione S-transferase 26 kDa 51 (EC 2.5.1.18) (GST51) (FH51)
DE (GST class-alpha).
OS Fasciola hepatica (Liver fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
OC Echinostomidae; Echinostomata; Fascioloidae; Fasciolidae; Fasciola.
OX NCBI_TaxID=6192;
RN [1];
RP SEQUENCE FROM N.A.
RA MEDLINE=92155306; PubMed=1740183;
RA Parascio M., Wilson L.R., Cramer S.L., Wijffels G.L., Spithill T.W.;
RT "Molecular characterization of cDNA sequences encoding glutathione S-
RL transferases of Fasciola hepatica.";
RN Exp. Parasitol. 74:232-237(1992).
RP [2];
RN ERRATUM.
RP MEDLINE=94039664; PubMed=8224094;
RA Parascio M., Wilson L.R., Cramer S.L., Wijffels G.L., Spithill T.W.;
RL Exp. Parasitol. 77:385-385(1993).
RN [3];
RP SEQUENCE OF 9-217 FROM N.A.
RA MEDLINE=93228188; PubMed=7682383;
RA Muro A., Rodriguez-Molina J.R., Hillyer G.V.;
RT "Sequence analysis of a Fasciola hepatica glutathione S-transferase
RL cDNA clone".;
RN Am. J. Trop. Med. Hyg. 48:457-463(1993).
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC PARASITE DETOXICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
CC HAEMATIN IN THE PARASITE GUT.
CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M7682; AAA29141.1; -.
DR HSSP; P31670; 1PHE.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam; PF00043; GST_C_1.
DR Pfam; PF02798; GST_N_1.
KW Transferase; Antigen; Multigene family.
FT INIT MET 0
FT CONFLICT 83 83 T -> S (IN REF. 3).
FT CONFLICT 134 134 N -> D (IN REF. 3).
FT CONFLICT 146 146 P -> T (IN REF. 3).
SQ SEQUENCE 217 AA; 25242 MW; EFCAB8FF7454EDC26 CRC64;

```

[illegible]

RA Mannervik B., Alin P., Gutenberg C., Jansson H., Tahlr M.K.,
 RA Marholm M., Joernvall H.;
 RT "Identification of three classes of cytosolic glutathione transferase
 RT common to several mammalian species: correlation between structural
 RT data and enzymatic properties.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:7202-7206(1985).
 RN [7]
 RP CHARACTERIZATION.
 RP STRAIN-CD-1; TISSUE-Liver;
 RX MEDLINE-96189427; PubMed-8605288;
 RA Mitchell A.E., Morin D., Lame M.W., Jones A.D.;
 RT "Purification, mass spectrometric characterization, and covalent
 RT modification of murine glutathione S-transferases";
 RL Chem. Res. Toxicol. 8:1054-1062(1995).
 CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- MASS SPECTROMETRY: MW-25838.4; MW_ERR-2; METHOD-Electrospray.
 CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC -----
 DR EMBL: J03952; AAA37747.1; -;
 DR EMBL: J04632; AAA37705.1; -;
 DR EMBL: L13448; -; NOT_ANNOTATED_CDS.
 DR EMBL: BC003822; AAH03822.1; -;
 DR PIR: A20831; A20831.
 DR PIR: A28946; A28946.
 DR PIR: I24735; I24735.
 DR PIR: A34159; A34159.
 DR PIR: S33860; S33860.
 DR HSSP: P04905; 2GST.
 DR SWISS-2DPAGE: P10649; MOOSE.
 DR MGD: MGI:95860; Gstm1.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR InterPro: IPR003081; GST_mu.
 DR Pfam: PF00043; GST_Cy_1.
 DR Pfam: PF02798; GST_N; 1.
 DR PRINTS: PRO1267; GSTRNSFRSEM.
 DR Transferase: Multigene family.
 DR INIT_MET 0
 DR FT 0
 DR SQUENCE 217 AA; 25839 MW; A1EE3938F590B829 CRC64;
 Query Match 42.18; Score 489.5; DB 1; Length 217;
 Best Local Similarity 44.48; Pred. No. 1.0e-34;
 Matches 91; Conservative 39; Mismatches 70; Indels 5; Gaps 1;
 QY 4 ILGYMKIGVOPRLLELEYEEHLYERDEG-----DKMNKKFELGLEFPNPLY 58
 DB 3 ILGWNNRGLHPRMLLETDSDYDEKRTMGAPDRSQMLNEKFKLGLDPPNLYL 62
 QY 59 IDGDVKTQSGMAITRYADKHNLGSGPKERAEISLMLEGAVALDIRYGVSRVSKDEFTL 118
 DB 63 IDGSHKKTQSNALRYLARHNLHLDGFEERIRADIENQVMOTRMOLIMCYNPDEKQ 122
 QY 119 KVDPLSLPEMLKMEEDRLSHKTYLNDGHTHPFMLYDALDVLVMDPKLDFPKLYS 178
 DB 123 KPEPLKTYIPKMKLYSEFLKRPFPADKVTYVDLAVYDIDQYRMFEPKCLDAFPNLRD 182
 QY 179 FKRIEAIPOIDKYLKSSKYIAMPL 203
 DB 183 FLARFELKKISAVMKSSRIATPI 207

RESULT 9
 GTM1_RAT
 ID GTM1_RAT STANDARD; PRT; 217 AA.
 AC P04905;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glutathione S-transferase Yb1 (EC 2.5.1.18) (Chain 3) (GST M1-1)
 DE (GST class-mu 1).
 GN GSTM1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_Taxid:10116;
 RN [1]
 RP SEQUENCE FROM N.A. (CLONE PGTR200).
 RX MEDLINE-86312882; PubMed-2875437;
 RA Lai H.-C.J., Grove G., Tu C.-P.D.;
 RT "Cloning and sequence analysis of a cDNA for a rat liver glutathione
 RT S-transferase Yb subunit";
 RL Nucleic Acids Res. 14:6101-6114(1986).
 RN [2]
 RP SEQUENCE FROM N.A. (CLONE PGTA/C44).
 RX MEDLINE-86033768; PubMed-3840477;
 RA Ding G.-J., Ding V.D.-H., Rodkey J.A., Bennett C.D., Lu A.Y.H.,
 RA Pickert C.B.;
 RT "Rat liver glutathione S-transferases. Nucleotide sequence analysis
 RT of a Yb1 cDNA clone and prediction of the complete amino acid
 RT sequence of the Yb1 subunit";
 RL J. Biol. Chem. 260:13268-13271(1985).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-86224097; PubMed-3011803;
 RA Ding G.-J., Ding V.D.-H., Rodkey J.A., Bennett C.D., Lu A.Y.H.,
 RA Pickert C.B.;
 RT "Rat liver glutathione S-transferases. DNA sequence analysis of a Yb2
 RT cDNA clone and regulation of the Yb1 and Yb2 mRNAs by
 RT phenobarbital";
 RL J. Biol. Chem. 261:7952-7957(1986).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-87308179; PubMed-3040722;
 RA Chang C., Saltzman A.G., Sorensen N.S., Hlipakka R.A., Liao S.;
 RT "Identification of glutathione S-transferase Yb1 mRNA as the
 RT androgen-repressed mRNA by cDNA cloning and sequence analysis.";
 RL J. Biol. Chem. 262:11901-11903(1987).
 RN [5]
 RP SEQUENCE OF 1-23.
 RP STRAIN-Wistar; TISSUE-Olfactory epithelium;
 RX MEDLINE-93277499; PubMed-8503873;
 RA Ben-Arie N., Khen M., Lancet D.;
 RT "Glutathione S-transferases in rat olfactory epithelium;
 RT purification, molecular properties and odorant biotransformation.";
 RL Biochem. J. 292:379-384(1993).
 RN [6]
 RP MUTAGENESIS OF CYS-86.
 RX MEDLINE-91354218; PubMed-1883338;
 RA Hsieh J.-C., Huang S.-C., Chen W.-L., Lai Y.-C., Tam M.F.;
 RT "Cysteine 86 is not needed for the enzymic activity of glutathione S-
 RT transferase 3-3.";
 RL Biochem. J. 278:293-297(1991).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RX MEDLINE-93041702; PubMed-1420139;
 RA Ji X., Zhang P., Armstrong R.N., Gilliland G.L.;
 RT "The three-dimensional structure of a glutathione S-transferase from
 RT the yb1 gene class. Structural analysis of the binary complex of
 RT isoenzyme 3-3 and glutathione at 2.2-A resolution.";
 RL Biochemistry 31:10169-10184(1992).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RA Fu J.-H., Rose J., Tam M.F., Wang B.-C.;
 RT "New crystal forms of a mu-class glutathione S-transferase from rat
 RT liver.";

FT	STRAND	67	70
FT	HELIX	72	82
FT	TURN	83	84
FT	HELIX	90	114
FT	TURN	115	115
FT	TURN	117	118
FT	HELIX	119	128
FT	TURN	129	129
FT	HELIX	130	141
FT	TURN	142	143
FT	TURN	154	154
FT	HELIX	155	169
FT	TURN	171	176
FT	HELIX	178	188
FT	TURN	189	189
FT	HELIX	191	196
FT	TURN	197	198
FT	TURN	200	201
FT	TURN	210	211
SO	SEQUENCE	217 AA;	25782 MW; 2ACEBD49DA785118 CRC64;
Query Match		42.1%;	Score 489.5; DB 1; Length 217;
Best Local Similarity		44.4%;	Pred. No. 1.8e-34;
Matches	91; Conservative	39; Mismatches	70; Indels 5; Gaps 1;
Qy	4	ILGTWKIKGLVQPRLLILEYLEKEEHLKFEDEC-----DKWRKKFELGLEFNLPY 58	
Db	3	ILGWNVNRGLTHPIRLLEITLTDSSYEKKRYAMGADPDYDRSQMLNEKFKLGIDFNLPLY 62	
Qy	59	IDGDVKLTQSNAAIIRYIADKHNMLGSGSPKERAIEISMLEGAVDIRGVSRIVASKDFETL 118	
Db	63	IDGSKRTIQSNAINMYRLTKRHHLCGETEERIRADIVENQVNDNRKQIMLCINDFEKO 122	
Qy	119	KVDFLSKLPLEMLKMFEDRLSHRTYINGDVTHPDFMLYDALDVLVMDPMLDAEPKLV 178	
Db	123	KPEFLKTIPEKMKKLTSEFLGKRPMPFAGDKVTVYDFLAYDILDQYHIFEPKCIDAFPNLMD 182	
Qy	179	EKKRTAEIPQIDKYLKSSKYIAMPL 203	
Db	183	FLAREGLTKRISAYMKSSRYLSTPI 207	
RESULT 10			
GTM4_HUMAN			
ID	GTM4_HUMAN	STANDARD;	PRT; 218 AA.
AC	Q03013.		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Glutathione S-transferase Mu 4 (EC 2.5.1.18) (GSTM4-4) (GTS-Mu2)		
GN	(GST class-mu 4).		
GN	GSTM4.		
OC	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE=93352467; PubMed=8349586;		
RA	Comstock K.E., Johnson K.J., Rittenberg D., Henner W.D.;		
RA	"Isolation and analysis of the gene and cDNA for a human Mu class		
RT	glutathione S-transferase, GSTM4.";		
RL	J. Biol. Chem. 268:16958-16965(1993).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE=93228631; PubMed=8471052;		
RA	Zhong S., Spurr N.K., Hayes J.D., Wolf C.R.;		
RA	"Deduced amino acid sequence, gene structure and chromosomal location		
RT	of a novel human class Mu glutathione S-transferase, GSTM4.";		
RL	Biochem. J. 291:41-50(1993).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RP	TISSUE=Lung;		

FT	INIT	MET	0	0
FT	STRAND		2	5
FT	STRAND		7	7
FT	TURN		11	12
FT	HELIX		13	22
FT	TURN		23	24
FT	STRAND		27	29
FT	STRAND		32	32
FT	STRAND		35	35
FT	TURN		38	39
FT	STRAND		41	41
FT	HELIX		43	49
FT	TURN		50	52
FT	STRAND		61	64
FT	TURN		65	66
FT	STRAND		67	70
FT	HELIX		72	81
FT	TURN		82	84
FT	HELIX		90	113
FT	TURN		114	115
FT	TURN		117	118
FT	HELIX		119	127
FT	TURN		128	129
FT	HELIX		130	141
FT	TURN		142	143
FT	STRAND		146	146
FT	TURN		147	147
FT	STRAND		148	149
FT	TURN		150	150
FT	STRAND		151	151
FT	HELIX		154	169
FT	HELIX		171	176
FT	HELIX		178	189
FT	HELIX		191	197
FT	TURN		198	198
FT	TURN		214	215
SO	SEQUENCE	217 AA:	25613 MW:	25603A909482CA39 CRC64:
Query Match				
Best Local Similarity 41.6%; Score 484.5; DB 1; Length 217;				
Matches 95; Conservative 35; Mismatches 77; Indels 5; Gaps 12				
QY	5	LGWMIKRLVQPTRIILLEYLEKYEHEHLYERDEG-----DKRNKKFEULEFEPNLPYTI	59	
DB	4	LGWMIKRLANISIRLLLEYTDSSEYKRYTMDADYDRSQALNEKFKLGIDFPNLPYLI	63	
QY	60	DGDVVLITQSMAIRRYADKHNMLGSGPKERAETISMLEGAVLDIRVGSRIAYSKDEPTELK	119	
DB	64	DGTHKTIQSNMILRIARKHNLCGSESEKEDQIREDLLENQFMQSRQALATCTIDPDEFKEL	123	
QY	120	VDFLSKLEPMLEKMFEDRLSHKTTYLNGDHVTHDFMLYALDVLVLYMDPNCIDAFKLYSF	179	
DB	124	PEYLOALEPEMLKLYSQFGKQPFWFGDKITFEYDFIAYDVLEENQVFEPSCLDAFPLKDF	183	
QY	180	KKRIEAFQIDKYLKSSKYIAMPLOGQATG	211	
DB	184	ISREGELEKISAYMKSSRFLLPRPVFTKMAVMG	215	
RESULT 14				
GTMU_CAVPO				
AC	GTMU_CAVPO	STANDARD:	PRT:	217 AA.
DT	01-AUG-1990 (Rel. 15, Created)			
DT	01-AUG-1990 (Rel. 15, Last sequence update)			
DT	01-NOV-1995 (Rel. 32, Last annotation update)			
DE	Glutathione S-transferase B (EC 2.5.1.18) (Gst B) (Gst class-mu).			
GN	GSTM1.			
OS	Cavia porcellus (Guinea pig).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Hystriocognathi; Cavillidae; Cavia.			
OX	NCBI_TaxID=10141;			
EN	[1]			

RP SEQUENCE
RC TISSUE=Liver;
RX MEDLINE=90239691; PubMed-2332413;
RA Kamei K., Oshino R., Hara S.;
RT "Amino acid sequence of glutathione S-transferase b from guinea pig
RT liver.";
RL J. Biochem. 107:111-117(1990).
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.
DR PIR: JX0095; JX0095.
DR HSSP: P04905; 2GST.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR InterPro: IPR003081; GST_mu.
DR Pfam: PF00043; GST_C; 1.
DR Pfam: PF02798; GST_N; 1.
DR PRINTS: PRO1267; GSTRNSFRASEM.
DR Transferrase, Multigene family.
KW SEQUENCE 217 AA; 25719 MW; D29F7951D4E9365E CRC64;

Query Match 41.4%; Score 481.5; DB 1; Length 217;
Best Local Similarity 45.1%; Pred. No. 8.6e-34;
Matches 92; Conservative 35; Mismatches 72; Indels 5; Gaps 1;

QY 5 LGYKIKKLVLPPTALLLEYLEKYEHLERDEG-----DKRRKKFELGLEFPMLPYTI 59
Db 4 LGYNIIRIRLLHPRIIRILLEYTNSGYEKERYNMGDADPYDRSQWLNEKFKLGIDFPLPYLI 63
QY 60 DGDVLTQSMAIRIRIAKKNHMLGSPERARISMLGAVLDIRGVSRIVASXQPEFLK 119
Db 64 DGTHTLTQSMALIRIAKKNHMLGCTEETETIRMDLLENQYMDIRNQLMCTSPDEQAK 123
QY 120 VDFLSKLPKEMLKMEFDRLSHKTYLNGDHVTHPEMLYDALDVLVLYMDPNCIDAFKLYSF 179
Db 124 AEFLEGIDDKMKKLFESQFGLKLPWFAGNKLTLYDVLADVLQYRLMLEPCLEAFPLKDF 183
QY 180 KKRIEAIPOIDKYLKSKSYIAMPL 203
Db 184 ISREFGLEKISSYMKSSRFLPKPL 207

RESULT 15
GTWU_CRILLO
ID GTWU_CRILLO STANDARD; PRI: 217 AA.
AC Q00285;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutathione S-transferase Y1 (EC 2.5.1.18) (Chin 3) (GST class-mu).
OS Citreulione longicaudatus (long-tailed hamster) (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Citreulinae;
OC Citreulius.
OX NCBI_TaxID=10030;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91031445; PubMed-2226437; Debatisse M., Butlin G.;
RA de Saint Vincent B.R., Hyrien O.;
RT "Complete identification of mu class glutathione S-transferase genes and an
RT adenylate deaminase gene in coformycin-resistant Chinese hamster
RT fibroblasts.";
RL Eur. J. Biochem. 193:19-24(1990).
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL: X57489; CAA40726.1; -
DR PIR: S13202; S13202.
DR HSSP: P04905; 2GST.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C; 1.
DR Pfam: PF02798; GST_N; 1.
DR PRINTS: PR01267; GSTRNSFRASEM.
DR Transferase; Multigene family.
KW INIT_MET 0 BY SIMILARITY
FT SEQUENCE 217 AA; 2568 MW; 34D8A8FBD8627C0 CRC64;
SQ

Query Match 40.9%; Score 476.5; DB 1; Length 217;
Best Local Similarity 43.4%; Pred. No. 2.3e-33;
Matches 89; Conservative 37; Mismatches 74; Indels 5; Gaps 1;

QY 4 ILGYWKIKGLVQPTRLLEYLEEKYEHLRYERDEG----DKWRNKKFEGLIEFPNLPY 58
| | | | : | | | | | : | | : | | : | | | | | | | |
DB 3 ILGYWNVNGLNPRLLEYLETSYEKKYTKWDAPSDRSQWLNKFKLGLDFPNLPYL 62
| | | | : | | | | | : | | : | | : | | | | | | | |
QY 59 IDGVKLTQSMALIRYIDKRNMLGSPKEREISMLGAVLDIRYGVSRVAYSKDFETL 118
| | | | : | | | | | : | | : | | : | | | | | | | |
DB 63 IDGSHKITQSNALIRYIARKHNLGCTEERIRVDIVENQAMDTRMQLIMLCYNPDPEKQ 122
| | | | : | | | | | : | | : | | : | | | | | | | |
QY 119 KVDPLSKPLKMPEDRLSHKTYLNGDHYTHPDMLYDALDVLVLYMDPMLADAPPKLYS 178
| | | | : | | | | | : | | : | | : | | | | | | | |
DB 123 KPEFLKTIPEKKKKYSEFLGKRPFAGDKVTLGFLAYDVLDOYOMFEKCLDPFPNLKD 182
| | | | : | | | | | : | | : | | : | | | | | | | |
QY 179 FKKRIEALPQIDKYLKSSKRYIAMPL 203
| | | | : | | | | | : | | : | | : | | | | | | | |
DB 183 FLAREFGLKKISAYMKTSRFLRRPI 207
| | | | : | | | | | : | | : | | : | | | | | | | |

Search completed: May 20, 2003, 11:34:59
Job time : 5.14571 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 10:12:24 ; Search time 15.0413 Seconds
(Without alignments)
2986.328 Million cell updates/sec

Title: US-10-081-408-5
Perfect score: 1164
Sequence: 1 MSPILGYWKIKGLVPTRL.....IAMPILGQWATFGCGDHPK 218

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP viirus:*
16: SP bacteriophage:*
17: SP archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1034	88.8	218	5	094745
2	726	62.4	218	5	025595
3	707	60.7	218	5	09X19
4	520.5	44.7	218	6	09N0V4
5	514.5	44.2	219	13	090MM9
6	508.5	43.7	218	11	08R516
7	508.5	43.7	223	5	097117
8	485.5	41.7	218	6	09TSM5
9	483	41.5	221	6	09BE80
10	482.5	41.5	218	6	09TSM4
11	480.5	41.3	218	11	091Y83
12	472.5	40.6	218	11	09MU21
13	468.5	40.2	219	5	027653
14	467.5	40.2	219	5	016058
15	464.5	39.9	218	11	09DD25
16	464.5	39.9	218	11	09DCE8

17	454.5	39.0	225	4	096HA3	096ha3 homo sapien
18	448.5	38.5	225	6	09BEA9	09bea9 macaca fusc
19	444.5	38.2	195	4	005465	005465 homo sapien
20	444.5	38.2	220	4	08WEL1	08wel1 homo sapien
21	442.5	38.0	232	11	09D5J8	09d5j8 mus musculu
22	440.5	37.8	225	11	0921B2	0921b2 rattus norv
23	415.5	35.7	188	6	09MZB4	09mzb4 capra hircu
24	385.5	33.1	219	5	09U582	09u582 psoroptes o
25	352	30.2	181	4	08TC98	08tc98 homo sapien
26	330.5	28.4	220	5	08T7E4	08t7e4 boophilus m
27	290.5	25.0	125	6	029583	029583 sus scrofa
28	264.5	22.7	208	13	09DDU5	09ddu5 brachydanio
29	262.5	22.6	208	5	002636	002636 brugia mala
30	254.5	21.9	208	5	027711	027711 onchocerca
31	249	21.4	210	4	000460	000460 homo sapien
32	249	21.4	210	4	015690	015690 homo sapien
33	243	20.9	180	5	09NHB2	09nhb2 mytilus edu
34	242	20.8	209	6	09TYT8	09tyt8 capra hircu
35	222.5	19.1	216	13	09W647	09w647 oncorhynch
36	217.5	18.7	82	4	09UE37	09ue37 homo sapien
37	202.5	17.4	221	11	09DCU1	09dcu1 mus musculu
38	200.5	17.2	226	11	09JTX3	09jtx3 rattus norv
39	197.5	16.9	222	6	09N206	09n206 ovis aries
40	197	16.9	210	5	09N4X8	09n4x8 caenorhabd
41	196.5	16.9	198	5	08T1J9	08t1j9 dictyostell
42	196.5	16.9	221	11	P70686	P70686 mesocricetu
43	192.5	16.5	208	5	P91505	P91505 caenorhabd
44	191.5	16.5	223	6	029057	029057 sus scrofa
45	190.5	16.4	222	6	09X530	09x530 ovis aries

ALIGNMENTS

RESULT 1	ID	PRELIMINARY;	PRT;	218 AA.
Q94745	094745			
AC	094745			
DT	01-FEB-1997 (TREMBLrel. 02, Created)			
DT	01-FEB-1997 (TREMBLrel. 02, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	26kD glutathione S-transferase.			
OS	Schistosoma mekongi.			
OC	Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoidida;			
CC	Schistosomatoidae; Schistosomatidae; Schistosoma.			
OX	NCBI_Taxid=38744;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=THAILAND;			
RX	MEDLINE=98222395; PubMed=9561610;			
RA	Grams S.V., Grams R., Korge G., Vinyant V., Upatham S.;			
RT	"Cloning and sequencing of the 26 kDa glutathione S-transferase gene			
RT	of Schistosoma mekongi."			
RL	Southeast Asian J. Trop. Med. Public Health 28:570-574(1997).			
DR	EMBL; Y07663; CA68944.1; -.			
DR	HSSP; P08515; 1GTA.			
DR	InterPro: IPR004046; GST_Cterm.			
DR	InterPro: IPR004045; GST_Nterm.			
DR	Pfam; PF00043; GST_C; 1.			
DR	Pfam; PF02796; GST_N; 1.			
KW	TRANSFERASE.			
SO	SEQUENCE 218 AA; 25486 MW; D380620B19DE23A CRC64;			
Query Match	88.8%; Score 1034; DB 5; Length 218;			
Best Local Similarity	89.0%; Pred. No. 1.1e-81;			
Matches 194; Conservative 11; Mismatches 13; Indels 0; Gaps 0;				
OY	1 MSPILGYWKIKGLVPTRLLEYLEEKHEEHLDEGDKNRKKEFELGFPNLPYID 60			
DB	1 MAPILGYWKIKGLVPTRLLEYLEEKHEEHLDEGDKNRKKEFELGFPNLPYID 60			
OY	61 GDVLTOSMAITRIYADKHNNLGGSPKERAEISMLEGAVLDIRGVSHIAYSKFETLKV 120			

Db	61	GDVKLTGSMATIRYIAOKAHNNLGGSGKEBAEITMLEGAVSPDIRSGVSRINAYNNDFELKV	120
Qy	121	DELKSLRPMELKMFEDRLSHKTYTLLKGGDHYTHHDFMLYALADVLYLMDMGCADAPKLYSEK	180
Db	121	DELKSLRPMELKMFEDRLCHKTYLNGDKNVTHHDFMLYALADVLYLMDKCDLAPKLYCFK	180
Qy	181	KRIEALPOIDKYLKSSKTYIANPLOGMQTREGGDDHPK	218
Db	181	KRIENIPQINELYLRSSKTYIENPLOGMQTREGGDDHPK	218

RESULT 2

ID	Q25595	PRELIMINARY;	PRT;	218	AA.
AC	Q25595.				
DT	01-NOV-1996	(TREMBLrel. 01, Created)			
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)			
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)			
DE	Putative glutathione transferase.				
OS	Clonorchis sinensis.				
OC	Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;				
OC	Opisthorchiida; Opisthorchata; Opisthorchiidae; Opisthorchidae;				
OC	Clonorchis.				
OX	NCBI_TaxID=79923;				

RP SEQUENCE FROM N.A.
RA Hong S.-J., Lee D.-H.:
RT "Cloning and over expression of 26 kDa glutathione S-transferase from
RT *Clonorchis sinensis*.";
RL Thesis (1996), Parastology, Chung-Ang University College of Medicine,
Secul.
[2]
RN SEQUENCE FROM N.A.

.RA Hong S.-J.;
 .RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases
 DR EMBL; L479922; AA846359.3; -
 DR HSSP; P31670; 1PHE.
 DR InterPro; IPR0044046; GST_Cterm.
 DR InterPro; IPR0044045; GST_Nterm.
 DR Pfam; PF00043; GST_C_1.
 DR Pfam; PF02798; GST_N_1.
 KM Transferrase.
 SQ SEQUENCE 218 AA: 25038 MW: 7CB17C7B837A0B7C CRC64;

Query Match	62.48;	Score	725;	DB	5;	Length	216;	
Best Local Similarity	59.28;	Pred. No.	4.8e-55;					
Matches 129;	Conservative	38;	Mismatches	51;	Indels	0;	Gaps	0;

[illegible]

RESULT 3

ID	NAME	PRELIMINARY	PRT	218 AA
09XYL9				
AC 09XYL9				
DT 01-NOV-1999	(TREMBLrel. 12, Created)			
DT 01-NOV-1999	(TREMBLrel. 12, Last sequence update)			
DT 01-JUN-2002	(TREMBLrel. 21, Last annotation update)			
DE	Glutathione S-transferase.			

GN GST-1.
OS Fasciola gigantica.
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
OC Echinostomida; Echinostomata; Fascioloidae; Fasciola
OX NCBI_TaxID=46835;

Sub
RT
RT

RA Grams S.V., Grams R., Sobhon P., Vivanant V., Upatham E.S.;
RT "Molecular cloning of expressed antigens from *Parascaris* sp.",
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ
DR EMBL; AF112567; AAD33997.1; -
DR HSSP; P31670; 1FHE.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF00043; GST_C; 1.
DR Pfam: PF02798; GST_N; 1.
KW Transferrase.
SQ SEQUENCE 218 AA; 25337 MW; E9423D75C3F2BEAF CR664;

Query Match	60.7%	Score 707	DB 5:	length 218;
Best Local Similarity	58.7%	Pred. No. 2.1e-53;		
Matches 125; Conservative	36;	Mismatches 32;	Indels 0;	Gaps 0

```

0Y 5 LGYMKIKLOVPTRIILLEYLEEKVEEHLYEYDEDDKRNKKFELGEPFNP.PYIIOGVK 64
Db 5 LGYMKIRLOQPVRLILEYLEDEEYEEHLYDGRDEKMKLGDKFNNGJLDLPNP.PYIIODKK 64
0Y 65 LTOSALIRYIADNHNHMGSSPKERALE.SMEGAVIIDIRGVSHIAAXSKOPELTKUDLS 124
Db 65 LTOSALIRYIADNHNHMGSSPKERALE.SMEGAVIIDIRGVSHIAAXSKOPELTKUDLS 124

```

[illegible]

RESULT 4

```

ID      Q9NOV4;      PRELIMINARY;      PRT;      218 AA.
AC      Q9NOV4;
DT      01-OCT-2000 (TREMBLrel. 15, Created)
DT      01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE      Class mu glutathione S-transferase.
GN      GSTM..
OS      Bos taurus (Bovine).
OC      Eumariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae
OC      Bovidae; Bovinae; Bos.
CX      NCBI_TaxID=9913;
[]
RN      SEQUENCE FROM N.A.
RP      TISSUE=LENS;
RA      Jimenez-Asensio J.V., Garland D.;
RT      "A lens glutathione S-transferase, class mu, with thiol-specific
RT      antioxidant activity ";
RL      Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
DR      EMBL: AF249588; AAF64308.1; -.
DR      HSSP; P09488; IGTU.
DR      InterPro: IPR004046; GST_Cterm.
DR      InterPro: IPR003081; GST_mu.
DR      InterPro: IPR004045; GST_Nterm.
DR      Pfam: PF00043; GST_C; 1.
DR      Pfam: PF02798; GST_N; 1.
DR      PRINTS: PR01267; GSTRNFSFASPM.
KW      Transferase.
SQ      SEQUENCE      218 AA; 25635 MW; 3D02EA0F43C07B0A CRC64;

```

Query Match 44.7%; Score 520.5; DB 6; Length 218;
 Best Local Similarity 47.6%; Pred. No. 2.9e-37;
 Matches 99; Conservative 36; Mismatches 68; Indels 5; Gaps 1;

QY 1 MSPILGWTKIKGLVOPRLLLEYLEEKEEHLRYERDEG-----DKRNKKFELGLEFPNL 55
 DB 1 MPMLIGWMDIRGLAHAIIRLLETTDTYERQYSGADPDYDRSQWLNKFKGLDPEPNL 60
 QY PYIDGDKVLTQSMALIRYIADKHNMLGSPKREAEISMLGAVLDIRYGSRIAYSKDF 115
 DB 61 PYLIDGTHKLTQSNAILIRYIARKHNMLGCEETEEMIRVDILENQWMDVRLAMARICSPDF 120
 QY 116 ETLKVDLSKLPEMLKMFEDRLSHKTYLNGDVTHTPDMLYDALDVLYMDPCLDAFPK 175
 DB 121 EKLKPGFLKEIKPEKIKLFSEFLGKRPWAGDKLTLYVDVLYDMDHRIPEPKCLDAFPN 180

QY 176 LVSEFKRIEALPOIDKYLKSSKYIAMPLOQWATFG 203
 DB 181 LKDFISRFEGLKISAYMKSSRFPLPPL 208

RESULT 5
 090MM9 PRELIMINARY; PRT; 219 AA.

AC 090MM9;
 DT 01-DEC-2001 (TREMBLrel, 19, Created)
 DT 01-DEC-2001 (TREMBLrel, 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel, 21, Last annotation update)
 DE Glutathione S-transferase (EC 2.5.1.18).
 GN GSTM1.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 OC Xenopodidae; Xenopus.
 NC NCBL_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA De Luca A., Favaloro B., Sacchetta P., Angelucci S., Di Ilio C.;
 RT "Molecular cloning, expression and site-directed mutagenesis of a
 RT 116er Mu-class glutathione S-transferase from Xenopus laevis.";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ116998; CAD01094.1; -
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF02798; GST_N; 1.
 DR Pfam: PF02798; GST_N; 1.
 KW Transferase.
 FT CHAIN 2 219 XLGSTM1.
 SQ SEQUENCE 219 AA; 25359 MW; 5F7FB8AEDDA82DFD CRC64;

Query Match 44.2%; Score 514.5; DB 13; Length 219;
 Best Local Similarity 47.7%; Pred. No. 9.8e-37;
 Matches 103; Conservative 34; Mismatches 74; Indels 5; Gaps 1;

QY 1 MSPILGWTKIKGLVOPRLLLEYLEEKEEHLRYERDEG-----DKRNKKFELGLEFPNL 55
 DB 1 MVLILGYMDIRGLAHAIIRLLETTDTYERQYSGADPDYDRSQWLNKFKGLDPEPNL 60
 QY 56 PYIDGDKVLTQSMALIRYIADKHNMLGSPKREAEISMLGAVLDIRYGSRIAYSKDF 115
 DB 61 PYLIDGTHKLTQSNAILIRYIARKHNMLGCEETEEMIRVDILENQWMDVRLAMARICSPDF 120
 QY 116 ETLKVDLSKLPEMLKMFEDRLSHKTYLNGDVTHTPDMLYDALDVLYMDPCLDAFPK 175
 DB 121 ETLKPGFLKEIKPEKIKLFSEFLGKRPWAGDKLTLYVDVLYDMDHRIPEPKCLDAFPN 180

QY 176 LVSEFKRIEALPOIDKYLKSSKYIAMPLOQWATFG 211
 DB 181 LQAFLTREALPAIDAYVKSFRFKTPIINRRASWG 216

RESULT 6
 08R516 PRELIMINARY; PRT; 218 AA.

QY 1 MSPILGWTKIKGLVOPRLLLEYLEEKEEHLRYERDEG-----DKRNKKFELGLEFPNL 55
 DB 1 MPMLIGWMDIRGLAHAIIRLLETTDTYERQYSGADPDYDRSQWLNKFKGLDPEPNL 60
 QY PYIDGDKVLTQSMALIRYIADKHNMLGSPKREAEISMLGAVLDIRYGSRIAYSKDF 115
 DB 61 PYLIDGTHKLTQSNAILIRYIARKHNMLGCEETEEMIRVDILENQWMDVRLAMARICSPDF 120
 QY 116 ETLKVDLSKLPEMLKMFEDRLSHKTYLNGDVTHTPDMLYDALDVLYMDPCLDAFPK 175
 DB 121 EKLKPGFLKEIKPEKIKLFSEFLGKRPWAGDKLTLYVDVLYDMDHRIPEPKCLDAFPN 180

QY 176 LVSEFKRIEALPOIDKYLKSSKYIAMPLOQWATFG 211
 DB 181 LKDFAREVELKRIKSAVYKTSRFLPTPLTKVATWG 216

RESULT 7
 097117 PRELIMINARY; PRT; 223 AA.

AC 097117;
 DT 01-MAY-1999 (TREMBLrel, 10, Created)
 DT 01-MAY-1999 (TREMBLrel, 10, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel, 21, Last annotation update)
 DE Glutathione S-transferase.
 OS Boophilus microplus (Cattle tick).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 OC Parasitiformes; Ixodida; Ixodidae; Boophilus.
 NC NCBL_TaxID=6941;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA He H., Chen A.C., Davey R.B., Ivie G.W., George J.E.;
 RT MEDLINE=99381232; PubMed=10451925;
 RT "Characterization and molecular cloning of a glutathione S-transferase
 RT gene from the tick, Boophilus microplus (Acari: Ixodidae).";
 RL Insect Biochem. Mol. Biol. 29:737-743(1999).
 DR EMBL: AF077609; AAD15991.1; -
 DR HSP: P20136; 1GSU.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR003081; GST_mu.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR PRINTS: PRO1267; GSTRNSFRASEM.
 KW Transferase.
 SQ SEQUENCE 223 AA; 25604 MW; F276ACE2D8259174 CRC64;

Query Match 43.7%; Score 508.5; DB 11; Length 218;
 Best Local Similarity 44.9%; Pred. No. 3.2e-36;
 Matches 97; Conservative 40; Mismatches 74; Indels 5; Gaps 1;

QY 1 MSPILGWTKIKGLVOPRLLLEYLEEKEEHLRYERDEG-----DKRNKKFELGLEFPNL 55
 DB 1 MPMLIGWMDIRGLAHAIIRLLETTDTYERQYSGADPDYDRSQWLNKFKGLDPEPNL 60
 QY PYIDGDKVLTQSMALIRYIADKHNMLGSPKREAEISMLGAVLDIRYGSRIAYSKDF 115
 DB 61 PYLIDGTHKLTQSNAILIRYIARKHNMLGCEETEEMIRVDILENQWMDVRLAMARICSPDF 120
 QY 116 ETLKVDLSKLPEMLKMFEDRLSHKTYLNGDVTHTPDMLYDALDVLYMDPCLDAFPK 175
 DB 121 EKLKPGFLKEIKPEKIKLFSEFLGKRPWAGDKLTLYVDVLYDMDHRIPEPKCLDAFPN 180

QY 176 LVSEFKRIEALPOIDKYLKSSKYIAMPLOQWATFG 211
 DB 181 LKDFAREVELKRIKSAVYKTSRFLPTPLTKVATWG 216

RESULT 7
 097117 PRELIMINARY; PRT; 223 AA.

AC 097117;
 DT 01-MAY-1999 (TREMBLrel, 10, Created)
 DT 01-MAY-1999 (TREMBLrel, 10, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel, 21, Last annotation update)
 DE Glutathione S-transferase.
 OS Boophilus microplus (Cattle tick).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 OC Parasitiformes; Ixodida; Ixodidae; Boophilus.
 NC NCBL_TaxID=6941;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA He H., Chen A.C., Davey R.B., Ivie G.W., George J.E.;
 RT MEDLINE=99381232; PubMed=10451925;
 RT "Characterization and molecular cloning of a glutathione S-transferase
 RT gene from the tick, Boophilus microplus (Acari: Ixodidae).";
 RL Insect Biochem. Mol. Biol. 29:737-743(1999).
 DR EMBL: AF077609; AAD15991.1; -
 DR HSP: P20136; 1GSU.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR003081; GST_mu.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR PRINTS: PRO1267; GSTRNSFRASEM.
 KW Transferase.
 SQ SEQUENCE 223 AA; 25604 MW; F276ACE2D8259174 CRC64;

Query Match 43.7%; Score 508.5; DB 5; Length 223;

RT liver.";
 RL Toxicol. Sci. 56:26-36(2000).
 DR EMBL: AF200710; AAF08540.1; .
 DR HSSP: P28161; 1HNA.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR003081; GST_mu.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR PRINTS: PR01267; GSTNSFRASEM.
 KM Transferase.
 SQ SEQUENCE 218 AA; 25708 MW; F2E509C3949F9051 CRC64;

Query Match 41.5%; Score 482.5; DB 6; Length 218;
 Best Local Similarity 44.4%; Pred. No. 5,7e-34;
 Matches 96; Conservative 35; Mismatches 80; Indels 5; Gaps 1;

OY 1 MSPILGYNKIGLVOPTRLLLEYLEKEEHLERDEG-----DKNRKKFELGLEFPNL 55
 DB 1 MPMILGYNWINGLASHIRLLLEYGSSYEKKYTMGDADYDRSQWMLNEKKEKLGDFPNL 60
 OY 56 PYIDGVKLTOSMAIIRYADKHNMLGSPKERAISMLEGAVLDIRYGSRIAYSKDF 115
 DB 61 PYLDGTHKRTOSNAIIRYARKHNLGCEETEKEKREDILENQLMDNMQALRCYDPDF 120
 OY 116 ETLAVDFLSKLPKMEKFEEDRLSHKTYLNGDVTHTPDMYDALDVLVMDPMCLDAFPK 175
 DB 121 EKLPEYLEGPEMLKTYSGFQKOPWLGDKTFEVDIAYDVLERNGVFERSCLDAFPN 180
 OY 176 LVSEKRIEAIPOIDKYLKSSKIYAMPLOQWQATFG 211
 DB 181 LKDFISREGLKISAYMKSSRFPRPVFTKMAVWG 216

RESULT 11
 091Y83 PRELIMINARY; PRT: 218 AA.

AC 091Y83;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Glutathione S-transferase subunit gyc.
 OS Cavia porcellus (Guinea pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Hystriocognathi; Caviidae; Cavia.
 OX NCBI_TaxID=10141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HARTLEY; TISSUE=LIVER;
 RX MEDLINE=98297271; PubMed=9633615;
 RA Hiratsuka A., Ogura K., Fujioke H., Sakamoto Y., Okuda H., Wada K.,
 Tanaka T., Nishiyama T., Watabe T.;
 RT "Guinea pig liver Mu-class glutathione S-transferase M1-2 cross-reacts
 with antilodopes to both rat Mu- and theta-class glutathione S-
 transferases.";
 RL Arch. Biochem. Biophys. 354:188-196(1998).
 DR EMBL: AB000488; BAB47185.1; .
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 KM Transferase.
 SQ SEQUENCE 218 AA; 25695 MW; D2B9E31F9FEFA1E8 CRC64;

Query Match 41.3%; Score 480.5; DB 11; Length 218;
 Best Local Similarity 44.7%; Pred. No. 8,5e-34;
 Matches 93; Conservative 34; Mismatches 76; Indels 5; Gaps 1;

OY 1 MSPILGYNKIGLVOPTRLLLEYLEKEEHLERDEG-----DKNRKKFELGLEFPNL 55
 DB 1 MPMILGYNWINGLASHIRLLLEYGSSYEKKYTMGDADYDRSQWMLNEKKEKLGDFPNL 60
 OY 56 PYIDGVKLTOSMAIIRYADKHNMLGSPKERAISMLEGAVLDIRYGSRIAYSKDF 115

DB 61 PYLDGTHKRTOSNAIIRYARKHNLGCEETEKEKREDILENQLMDNMQALRCYDPDF 120
 OY 116 ETLAVDFLSKLPKMEKFEEDRLSHKTYLNGDVTHTPDMYDALDVLVMDPMCLDAFPK 175
 DB 121 EKLPEYLEGPEMLKTYSGFQKOPWLGDKTFEVDIAYDVLERNGVFERSCLDAFPN 180
 OY 176 LVSEKRIEAIPOIDKYLKSSKIYAMPLOQWQATFG 211
 DB 181 LKDFISREGLKISAYMKSSRFPRPVFTKMAVWG 216

RESULT 12
 09WU21 PRELIMINARY; PRT: 218 AA.

AC 09WU21;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Glutathione S-transferase Yb4 (Ec 2.5.1.18).
 GN GSTYB4.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY;
 RX MEDLINE=88298790; PubMed=3403534;
 RA Lai H.C., Qian B., Grove G., Tu C.P.;
 RT "Gene expression of rat glutathione S-transferases. Evidence for gene
 conversion in the evolution of the Yb multigene family.";
 RT J. Biol. Chem. 263:11389-11395(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY;
 RA Qian B., Tu C.P.;
 RT "The Rat Glutathione S-Transferase Yb4 genomic clone sequence.";
 RL Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF106661; AAD22630.1; .
 DR HSSP: P04905; 2GST.
 DR InterPro: IPR004046; GST_Cterm.
 DR InterPro: IPR004045; GST_Nterm.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 KM Transferase.
 SQ SEQUENCE 218 AA; 25643 MW; 4E9C22F9AEEAC6C4 CRC64;

Query Match 40.6%; Score 472.5; DB 11; Length 218;
 Best Local Similarity 43.3%; Pred. No. 4,2e-33;
 Matches 90; Conservative 36; Mismatches 77; Indels 5; Gaps 1;

OY 1 MSPILGYNKIGLVOPTRLLLEYLEKEEHLERDEG-----DKNRKKFELGLEFPNL 55
 DB 1 NAMILGYNVRLTPRILLEYLTDSTNVEKKRYVWGADAPNFRSQWMLSEKFLGLDIFNL 60
 OY 56 PYIDGVKLTOSMAIIRYADKHNMLGSPKERAISMLEGAVLDIRYGSRIAYSKDF 115
 DB 61 PYLDGSHKVTOSNAIIRYARKHNLGCEETEKEKREDILENQLMDNMQALRCYDPDF 120
 OY 116 ETLAVDFLSKLPKMEKFEEDRLSHKTYLNGDVTHTPDMYDALDVLVMDPMCLDAFPK 175
 DB 121 EKQKPEFLKSIPEKKKIYSEFLGKRPRWAGDVTYVDFLAYDLDIYRMEFECDAPFN 180
 OY 176 LVSEKRIEAIPOIDKYLKSSKIYAMPLOQWQATFG 211
 DB 181 LKDFISREGLKISAYMKSSRFPRPVFTKMAVWG 216

RESULT 13
 027653 PRELIMINARY; PRT: 219 AA.
 ID 027653;
 AC 027653;


```
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Glutathione transferase (EC 2.5.1.18).
CN EMGST1.
OS Echinococcus multilocularis.
OC Eukaryota; Metazoa; Platyhelminthes; Cestoda; Eucestoda;
OC Cyclophyllidae; Taeniidae; Echinococcus.
OX NCBI_Taxid=6211;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96379220; Pubmed=9784771;
RA Liebau E., Muller V., Lucius R., Walter R.D., Henkle-Duhrsen K.;
RT "Molecular cloning, expression and characterization of a recombinant
RT glutathione S-transferase from Echinococcus multilocularis.";
RL Mol. Biochem. Parasitol. 77:49-56(1996).
DR EMBL; X85736; CA59739.1; -.
DR HSSP; P20136; 1GSU.
DR InterPro; IPR004046; GST_Cterm.
DR Pfam; PF00043; GST_C; 1.
DR Pfam; PF02798; GST_N; 1.
DR Transferrase.
SO SEQUENCE 219 AA; 25470 MW; 7AFAB20AC56F745B CRC64;

Query Match 40.2%; Score 468.5; DB 5; Length 219;
Best Local Similarity 41.9%; Pred. No. 9,4e-33;
Matches 91; Conservative 41; Mismatches 80; Indels 5; Gaps 1;

OY 1 MSPILGKIKIGVOPTRLLLEYEKEEHLYE-----RDGDKWRKKFELGLEFPNL 55
DB 1 MAPTLAYWDINGLAEQSLRLKYLEVEYDDKRYKIGSAPTRDSAMLSEKESGLDGFNL 60

OY 56 PYYIDGVKLTOSMAIIRYIADKHNMLGSPKERAIEISMLEGAVDIRGVSRINYSKDF 115
DB 61 PYYIDGVKLTOSMAIIRYIADKHNMLGSPKERAIEISMLEGAVDIRGVSRINYSKDF 120

OY 116 ETLKVDLSKLPENLKMEDRLSHKTYLNGDVHPDMLYDALDVLVYMPMCLDAEPK 175
DB 121 ETLKVDLSKLPENLKMEDRLSHKTYLNGDVHPDMLYDALDVLVYMPMCLDAEPK 180

OY 176 LVSEFKRIEAIPOIDKYLKSSRYIAMPLOGQATPGG 212
DB 181 LKAYLSRENPALRDYASKKFTKPCNGASAKWRG 217

RESULT 14
O16058 PRELIMINARY; PRT; 219 AA.
AC O16058;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Glutathione S-transferase.
CN GSTWU.
OS Echinococcus granulosus.
OC Eukaryota; Metazoa; Platyhelminthes; Cestoda; Eucestoda;
OC Cyclophyllidae; Taeniidae; Echinococcus.
OX NCBI_Taxid=6210;
RN [1]
RP SEQUENCE FROM N.A.
RA Fernandez V., Zaha A., Fernandez C.;
RL Submitted (May-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Fernandez Mancebo V., Chalar C., Martinez C., Zaha A.,
RA Fernandez Granja C.;
RT "ggGST: a glutathione S-transferase gene from Echinococcus
RT granulosus.";
RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF005928; AAB66318.1; -.
DR EMBL; AF101269; AADI6438.1; -.
DR HSSP; P20136; 1GSU.
```

```
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR004045; GST_Nterm.
DR Pfam; PF00043; GST_C; 1.
DR Pfam; PF02798; GST_N; 1.
DR Transferrase.
SO SEQUENCE 219 AA; 25553 MW; 8883E70AD075D154 CRC64;

Query Match 40.2%; Score 467.5; DB 5; Length 219;
Best Local Similarity 41.9%; Pred. No. 1,1e-32;
Matches 91; Conservative 41; Mismatches 80; Indels 5; Gaps 1;

OY 1 MSPILGKIKIGVOPTRLLLEYEKEEHLYE-----RDGDKWRKKFELGLEFPNL 55
DB 1 MAPTLAYWDINGLAEQSLRLKYLEVEYDDKRYKIGSAPTRDSAMLSEKESGLDGFNL 60

OY 56 PYYIDGVKLTOSMAIIRYIADKHNMLGSPKERAIEISMLEGAVDIRGVSRINYSKDF 115
DB 61 PYYIDGVKLTOSMAIIRYIADKHNMLGSPKERAIEISMLEGAVDIRGVSRINYSKDF 120

OY 116 ETLKVDLSKLPENLKMEDRLSHKTYLNGDVHPDMLYDALDVLVYMPMCLDAEPK 175
DB 121 ETLKVDLSKLPENLKMEDRLSHKTYLNGDVHPDMLYDALDVLVYMPMCLDAEPK 180

OY 176 LVSEFKRIEAIPOIDKYLKSSRYIAMPLOGQATPGG 212
DB 181 LKAYLSRENPALRDYASKKFTKPCNGASAKWRG 217

RESULT 15
O9DD25 PRELIMINARY; PRT; 218 AA.
AC O9DD25;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE 0610005A07Rik protein.
CN RGRIP1 OR 0610005A07Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
RX MEDLINE=21085660; Pubmed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staib H., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carlini P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzatelli J., Mombaerts P.,
RA Nordone P., Ring B., Rothenbach C., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszewski A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK002213; BAB21939.1; -.
DR HSSP; P28161; 2GTU.
DR MGD; MGI:1915562; 0610005A07Rik.
DR MGD; MGI:1932134; Rgrip1.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR003081; GST_mu.
DR InterPro; IPR004045; GST_Nterm.
DR Pfam; PF00043; GST_C; 1.
```

DR Pfam: PF02798; GST_N; 1.
DR PRINTS; PR01267; GSTRNSFRASEM.
SQ SEQUENCE 218 AA; 25679 MW; 13471C2E08ABC115 CRC64;

Query Match 39.98; Score 464.5; DB 11; Length 218;
Best Local Similarity 41.28; Pred. No. 2.1e-32;
Matches 89; Conservative 42; Mismatches 80; Indels 5; Gaps 1;

```
QY 1 MSPILGYWKIKGLVQPTRLLEYLEEKYEHLVERDEG----DKWRKKFELGLEPPNL 55
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 MPMTLGIVWDIRGLAHAIRLEFLETDSYEERKRTMGDAPDYDOSQWLNEKFKLGIDPPNL 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 56 PYYIDGDKVLTQSMATIRYIADKHNMLGSPKERAELSMLEGAVIDIRYGVSRIVYSKDF 115
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 PYLIDGSHKITQSNALIRYIGRKHNLCGETEVERIRYVDILENQLMDNRVILARLCYNADF 120
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 116 ETIAKVDPLSKLPKMLKMFEDRLSHKITYLNGDHTHPDPMYDALDVLVLYMDPMCLDAPPK 175
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 EKLKPGYLEQLPGMMRLYSEFLGKRPWFAGDKITFVDIAYDVLERNQVEAKCLDAFPN 180
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 176 LVSEFKRRIEAIPOIDKYLKSSKYIAMPLOGMQATFG 211
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 LKDFIARFEGLKISDYMKTSRFLPRPMTKMATWG 216
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: May 20, 2003, 11:37:47
Job time : 16.0413 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 07:58:08 ; Search time 0.506582 seconds

(Without alignments)
1578.232 Million cell updates/sec

Title: US-10-081-408-6

Perfect score: 30

Sequence: 1 EALFQG 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq.101002.*

1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*

2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*

3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*

4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*

5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*

6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*

7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*

8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*

9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*

10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*

11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*

12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*

13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*

14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*

15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*

16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*

17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*

18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*

19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*

20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*

21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*

22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	100.0	11	AA1980	Sequence encoded b
2	30	100.0	11	AA1981	C-terminus of pol
3	30	100.0	26	AA1982	Sequence encoded b
4	30	100.0	26	AA1983	N-terminus of fusi
5	30	100.0	31	AA1984	Sequence encoded b
6	30	100.0	31	AA1985	N-terminus of fusi
7	30	100.0	31	AA1986	N-terminus of fusi
8	30	100.0	31	AA1987	N-terminus of fusi
9	30	100.0	989	AA1988	Herbicide-activ
10	30	100.0	2185	AA1989	Enteroviral poly

11	30	100.0	2206	13	AA1990	True type 3 polio
12	30	100.0	2209	3	AA1991	Sequence encoded b
13	28	93.3	168	21	AA1992	Sheep erythropoiet
14	28	93.3	2164	9	AA1993	Peptides translat
15	28	93.3	2164	9	AA1994	Sequence of the vi
16	27	90.0	24	22	AA1995	Smad interacting p
17	27	90.0	76	20	AA1996	Conopeptide prop
18	27	90.0	133	22	AA1997	Drosophila melanog
19	27	90.0	333	11	AA1998	Tumour-associated
20	27	90.0	365	21	AA1999	Human FAST-1 prote
21	27	90.0	365	23	AA2000	Human protein #1 r
22	27	90.0	533	22	AA2001	Peptide #5077 enco
23	27	90.0	631	22	AA2002	Enterococcus faeca
24	27	90.0	656	23	AA2003	Lactococcus lactis
25	26	86.7	8	22	AA2004	Presclasion protea
26	26	86.7	8	22	AA2005	Human Asp2 prestis
27	26	86.7	8	22	AA2006	Synthetic prestis
28	26	86.7	8	22	AA2007	Recognition site f
29	26	86.7	8	22	AA2008	Human Aspartyl pro
30	26	86.7	8	22	AA2009	Human rhinovirus 3
31	26	86.7	8	22	AA2010	Amino acid sequenc
32	26	86.7	8	22	AA2011	Human Aspd related
33	26	86.7	8	23	AA2012	Human rhinovirus 3
34	26	86.7	8	23	AA2013	Peptide r1 from th
35	26	86.7	32	16	AA2014	E.coli iss protein
36	26	86.7	113	21	AA2015	Avian E coli isola
37	26	86.7	113	22	AA2016	Human anti-Rh(D) a
38	26	86.7	127	22	AA2017	Human anti-Rh(D) a
39	26	86.7	183	22	AA2018	Drosophila melanog
40	26	86.7	199	20	AA2019	Chlamydia pneumoni
41	26	86.7	199	20	AA2020	Pseudomonas derugi
42	26	86.7	201	22	AA2021	Human novel secret
43	26	86.7	202	22	AA2022	D. maxinus thrombi
44	26	86.7	344	21	AA2023	Mammalian IBP-ALS
45	26	86.7	353	20	AA2024	

ALIGNMENTS

RESULT 1	AA1980	standard; protein; 11 AA.
ID	AA1980	
AC	AA1980	
DT	25-SEP-1993	(first entry)
XX		
DE	Sequence encoded by recombinant poliovirus at 3' end of	
DE	exogenous sequence.	
XX		
KW	Recombinant virus; proteolytic cleavage site; vaccine;	
KW	exogenous nucleic acid; replication competent.	
XX		
OS	Synthetic.	
XX		
FT	Key	Location/Qualifiers
FT	Region	1..7
FT	Cleavage-site	8
FT	Protein	/note="artificial"
FT		9..11
FT		/label= Polypeptide (PO)
PN	W09311251-A.	
XX		
PD	10-JUN-1993.	
XX		
PF	04-DEC-1992;	92WO-US10543.
XX		
PR	06-DEC-1991;	91US-0804893.
PR	18-SEP-1992;	92US-0947790.
XX		

PA (AMCY) AMERICAN CYANAMID CO.
 PA (WHED) WHITEHEAD INST BIOMEDICAL.
 XX
 XX Andino R, Feinberg M, Reilly PA, Weeks-Levy CL;
 DR MPI; 1993-197068/24.
 DR N-PSDB; AAQ43325.
 XX
 PT Recombinant viruses comprising artificial proteolytic cleavage
 PT site - useful as vaccines against bacterial, viral and fungal
 PT infections, parasitic diseases, cancer and allergies
 XX
 PS Disclosure; Fig 1A: 103pp; English.
 XX
 CC There are a number of locations within the poliovirus genome at
 CC which the exogenous nucleic acid sequence encoding the exogenous
 CC polypeptide and the nucleic acid sequences encoding the artificial
 CC proteolytic cleavage sites can be positioned to produce replication-
 CC competent recombinant polioviruses that express the encoded product.
 CC These sites within the genome of the poliovirus include a terminal
 CC end, the junction between the VP1 coding region and the 2A coding
 CC region, the junction between the 2A coding region and the 2B coding
 CC region and the junction between the 2C coding region and the 3A
 CC coding region (see AAQ43394-98). AAQ43324 and AAQ43325 show poliovirus
 CC genome modifications which permit insertion of exogenous nucleic
 CC acid sequences at an end of the polioviral genome.
 XX
 SQ Sequence 11 AA;
 Query Match 100.0%; Score 30; DB 14; Length 11;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 EALFQG 6
 Db 4 EALFQG 9
 RESULT 2
 ID AAY44185
 AC AAY44185 standard; peptide: 11 AA.
 XX
 AC AAY44185;
 XX
 DT 01-FEB-2000 (first entry)
 XX
 DE C-terminus of poliovirus protein for expression of exogenous genes.
 XX
 KW Antimicrobial; cytostatic; antiallergic; replication; recombinant;
 KW poliovirus; vector; expression; exogenous; protease; cleavage site;
 KW genome; primer; PCR; amplification; vaccine; bacterium; fungus; yeast;
 KW infection; parasitic disease; cancer; allergy.
 XX
 OS Synthetic.
 OS
 PN US5965124-A.
 PN
 PD 12-OCT-1999.
 PD
 PF 31-JAN-1995; 95US-0381637.
 PF
 PR 08-DEC-1992; 92US-0986729.
 PR 06-DEC-1991; 91US-0804893.
 PR 18-SEP-1992; 92US-0947790.
 PR
 PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
 PA (AMCY) AMERICAN CYANAMID CO.
 XX
 PI Feinberg M, Weeks-Levy CL, Reilly PA, Andino R;
 XX
 DR MPI; 1999-633429/54.
 DR N-PSDB; AAZ30653.
 XX

PT Replication-competent recombinant poliovirus for use as vaccine against
 PT bacterial, viral, fungal, and yeast infections, parasitic diseases,
 PT cancer and allergies -
 XX
 PS Disclosure; Fig 1A: 40pp; English.
 XX
 CC The invention relates to a replication-competent recombinant poliovirus
 CC vector which is able to express an exogenous polypeptide as a component
 CC of a recombinant polyprotein precursor. The precursor molecule also
 CC contains a protease cleavage allowing the precursor to be subsequently
 CC proteolytically processed by the polio protease 3C to release the
 CC exogenous protein. Preferably the exogenous sequence and the cleavage
 CC site sequence are inserted into the vector sequence at a location in
 CC the genome of the parent poliovirus such that they do not disrupt a
 CC polioviral sequence necessary for polioviral replication. This sequence
 CC corresponds to the artificial proteolytic site which is fused to the 3'
 CC end of the exogenous gene between the exogenous protein and the fusion
 CC polypeptide from the poliovirus. The replication-competent recombinant
 CC viruses are useful as vaccines against bacterial, viral, fungal, and
 CC yeast infections, parasitic diseases, cancer and allergies.
 XX
 SQ Sequence 11 AA;
 Query Match 100.0%; Score 30; DB 20; Length 11;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 EALFQG 6
 Db 4 EALFQG 9
 RESULT 3
 ID AAR37655
 AC AAR37655 standard; Protein: 26 AA.
 XX
 AC AAR37655;
 XX
 DT 25-SEP-1993 (first entry)
 XX
 DE Sequence encoded by recombinant poliovirus pMOV 3.1 (2C/3A).
 XX
 KW Recombinant virus; proteolytic cleavage site; vaccine;
 KW exogenous nucleic acid; replication competent.
 XX
 OS Synthetic.
 OS
 FH Key
 FT Cleavage-site 5..6
 FT /label- 3Cpro
 FT Cleavage-site 21..22
 FT /label- 3Cpro
 XX
 PN W09311251-A.
 PN
 PD 10-JUN-1993.
 PD
 PF 04-DEC-1992; 92WO-US10543.
 PF
 PR 06-DEC-1991; 91US-0804893.
 PR 18-SEP-1992; 92US-0947790.
 PR
 PA (AMCY) AMERICAN CYANAMID CO.
 PA (WHED) WHITEHEAD INST BIOMEDICAL.
 XX
 PI Andino R, Feinberg M, Reilly PA, Weeks-Levy CL;
 XX
 DR MPI; 1993-197068/24.
 DR N-PSDB; AAQ43297.
 XX
 PT Recombinant viruses comprising artificial proteolytic cleavage
 PT site - useful as vaccines against bacterial, viral and fungal
 PT infections, parasitic diseases, cancer and allergies

XX Disclosure; Fig 2B; 103pp; English.
 PS There are a number of locations within the poliovirus genome at
 CC which the exogenous nucleic acid sequence encoding the exogenous
 CC polypeptide and the nucleic acid sequences encoding the artificial
 CC proteolytic cleavage sites can be positioned to produce replication-
 CC competent recombinant polioviruses that express the encoded product.
 CC These sites within the genome of the poliovirus include a terminal
 CC end, the junction between the VP1 coding region and the 2A coding
 CC region, the junction between the 2A coding region and the 2B coding
 CC region and the junction between the 2C coding region and the 3A
 CC coding region (see AA043294-98).
 XX
 SQ Sequence 26 AA;
 Query Match 100.0%; Score 30; DB 14; Length 26;
 Best Local Similarity 100.0%; Pred. No. 5.3;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 EALFQG 6
 |||||
 Db 1 EALFQG 6
 RESULT 4
 ID AAY44190 standard; peptide; 26 AA.
 XX AAY44190;
 AC AAY44190;
 XX 01-FEB-2000 (first entry)
 DT
 XX N-terminus of fusion protein from modified poliovirus vector pMOV3.1.
 DE
 XX Antimicrobial; cytostatic; antiallergic; replication; recombinant;
 KM poliovirus; vector; expression; exogenous; protease; cleavage site;
 KM genome; primer; PCR; amplification; vaccine; bacterium; fungus; yeast;
 KM infection; parasitic disease; cancer; allergy.
 XX
 OS Poliovirus.
 OS Synthetic.
 OS
 XX US5965124-A.
 PN
 XX 12-OCT-1999.
 PD
 XX 31-JAN-1995; 95US-0381637.
 PF
 XX 08-DEC-1992; 92US-0986729.
 PR 06-DEC-1991; 91US-0804893.
 PR 18-SEP-1992; 92US-0947790.
 XX
 PA (MHED) WHITEHEAD INST BIOMEDICAL RES.
 PA (AMCY) AMERICAN CYANAMID CO.
 XX
 PI Feinberg M, Weeks-Levy CL, Reilly PA, Andino R;
 PI
 DR WPI; 1999-632429/54.
 DR N-PSDB; AA230658.
 XX
 XX Replication-competent recombinant poliovirus for use as vaccine against
 PT bacterial, viral, fungal, and yeast infections, parasitic diseases,
 PT cancer and allergies
 XX
 PS Example 2; Fig 2C; 40pp; English.
 XX
 CC The invention relates to a replication-competent recombinant poliovirus
 CC vector which is able to express an exogenous polypeptide as a component
 CC of a recombinant polypeptide precursor. The precursor molecule also
 CC contains a protease cleavage allowing the precursor to be subsequently
 CC proteolytically processed by the polio protease 3C to release the
 CC exogenous protein. Preferably the exogenous sequence and the cleavage

CC site sequence are inserted into the vector sequence at a location in
 CC the genome of the parent poliovirus such that they do not disrupt a
 CC polioviral sequence necessary for polioviral replication. This sequence
 CC corresponds to the N-terminal region of the fusion protein encoded by
 CC the multicloning site of the poliovirus vector pMOV3.1. The sequence
 CC spans the unique initiation site, the multicloning site, the protease C
 CC cleavage site and part of the N-terminus of the fusion polypeptide.
 CC The replication-competent recombinant viruses are useful as vaccines
 CC against bacterial, viral, fungal, and yeast infections, parasitic
 CC diseases, cancer and allergies.
 XX
 SQ Sequence 26 AA;
 Query Match 100.0%; Score 30; DB 20; Length 26;
 Best Local Similarity 100.0%; Pred. No. 5.3;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 EALFQG 6
 |||||
 Db 1 EALFQG 6
 RESULT 5
 ID AAR37652 standard; Protein; 31 AA.
 XX AAR37652
 AC AAR37652;
 XX 25-SEP-1993 (first entry)
 DT
 XX Sequence encoded by recombinant poliovirus pMOV 1.3 (NH2-terminal).
 DE
 XX Recombinant virus; proteolytic cleavage site; vaccine;
 KM exogenous nucleic acid; replication competent.
 KM
 XX Synthetic.
 OS
 OS Key Location/Qualifiers
 FT 16..21
 FT /label= poly-Gly
 FT 26..27
 FT /label= 3c pro
 XX
 PN WO9311251-A.
 XX
 XX 10-JUN-1993.
 PD
 XX 04-DEC-1992; 92WO-US10543.
 PF
 XX 06-DEC-1991; 91US-0804893.
 PR 18-SEP-1992; 92US-0947790.
 XX
 PA (AMCY) AMERICAN CYANAMID CO.
 PA (MHED) WHITEHEAD INST BIOMEDICAL.
 XX
 PI Andino R, Feinberg M, Reilly PA, Weeks-Levy CL;
 PI
 DR WPI; 1993-197068/24.
 DR N-PSDB; AA043294.
 XX
 XX Recombinant viruses comprising artificial proteolytic cleavage
 PT site - useful as vaccines against bacterial, viral and fungal
 PT infections, parasitic diseases, cancer and allergies
 XX
 PS Disclosure; Fig 2B; 103pp; English.
 XX
 CC There are a number of locations within the poliovirus genome at
 CC which the exogenous nucleic acid sequence encoding the exogenous
 CC polypeptide and the nucleic acid sequences encoding the artificial
 CC proteolytic cleavage sites can be positioned to produce replication-
 CC competent recombinant polioviruses that express the encoded product.
 CC These sites within the genome of the poliovirus include a terminal
 CC end, the junction between the VP1 coding region and the 2A coding

CC region, the junction between the 2A coding region and the 2B coding
 CC region and the junction between the 2C coding region and the 3A
 CC coding region (see AAQ43294-98).

XX Sequence 31 AA;

Query Match 100.0%; Score 30; DB 14; Length 31;
 Best Local Similarity 100.0%; Pred. No. 6.3;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 |||||
 Db 22 EALFOG 27

RESULT 6
 AAR37654
 ID AAR37654 standard; Protein; 31 AA.

XX AAR37654;
 AC AAR37654;
 DT 25-SEP-1993 (first entry)

XX Sequence encoded by recombinant poliovirus PMOV 2.5 (2A/2B).

XX Recombinant virus; proteolytic cleavage site; vaccine;
 KM exogenous nucleic acid; replication competent.

XX Synthetic.

XX Key Location/Qualifiers

FT Cleavage-site 5..6 /label- 3Cpro

FT Cleavage-site 26..27 /label- 3Cpro

XX MO9311251-A.

XX 10-JUN-1993.

XX 04-DEC-1992; 92WO-US10543.

XX 06-DEC-1991; 91US-0804893.

XX 18-SEP-1992; 92US-0947790.

XX (AMCY) AMERICAN CYANAMID CO.

XX (WHED) WHITEHEAD INST BIOMEDICAL.

XX Andino R, Feinberg M, Reilly PA, Weeks-Levy CL;

XX WPI; 1993-197068/24.

XX N-PSDB; AAQ43296.

XX Recombinant viruses comprising artificial proteolytic cleavage
 PT site - useful as vaccines against bacterial, viral and fungal
 PT infections, parasitic diseases, cancer and allergies

XX Disclosure; Fig 2B; 103pp; English.

XX There are a number of locations within the poliovirus genome at
 CC which the exogenous nucleic acid sequence encoding the exogenous
 CC polypeptide and the nucleic acid sequences encoding the artificial
 CC proteolytic cleavage sites can be positioned to produce replication-
 CC competent recombinant polioviruses that express the encoded product.
 CC These sites within the genome of the poliovirus include a terminal
 CC end, the junction between the VP1 coding region and the 2A coding
 CC region, the junction between the 2A coding region and the 2B coding
 CC coding region (see AAQ43294-98).

XX Sequence 31 AA;

Query Match 100.0%; Score 30; DB 14; Length 31;

Best Local Similarity 100.0%; Pred. No. 6.3;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 |||||
 Db 22 EALFOG 27

RESULT 7
 AAY44186
 ID AAY44186 standard; peptide; 31 AA.

XX AAY44186;
 AC AAY44186;
 DT 01-FEB-2000 (first entry)

XX N-terminus of fusion protein from modified poliovirus vector PMOV1.3.

XX Antimicrobial; cytostatic; antiallergic; replication; recombinant;
 KM poliovirus; vector; expression; exogenous; protease; cleavage site;
 KM genome; primer; PCR; amplification; vaccine; bacterium; fungus; yeast;
 KM infection; parasitic disease; cancer; allergy.

XX Poliovirus.
 OS Synthetic.

XX US5965124-A.

XX 12-OCT-1999.

XX 31-JAN-1995; 95US-0381637.

XX 08-DEC-1992; 92US-0986729.

XX 06-DEC-1991; 91US-0804893.

XX 18-SEP-1992; 92US-0947790.

XX (WHED) WHITEHEAD INST BIOMEDICAL RES.

XX (AMCY) AMERICAN CYANAMID CO.

XX Feinberg M, Weeks-Levy CL, Reilly PA, Andino R;

XX WPI; 1999-632429/54.

XX N-PSDB; AA230634.

XX Replication-competent recombinant poliovirus for use as vaccine against
 PT bacterial, viral, fungal, and yeast infections, parasitic diseases,
 PT cancer and allergies -

XX Example 2; Fig 2B; 40pp; English.

XX The invention relates to a replication-competent recombinant poliovirus
 CC vector which is able to express an exogenous polypeptide as a component
 CC of a recombinant polypeptide precursor. The precursor molecule also
 CC contains a protease cleavage allowing the polio protease 3C to release the
 CC proteolytically processed by the polio protease 3C to release the
 CC exogenous protein. Preferably the exogenous sequence and the cleavage
 CC site sequence are inserted into the vector sequence at a location in
 CC the genome of the parent poliovirus such that they do not disrupt a
 CC polioviral sequence necessary for polioviral replication. This sequence
 CC corresponds to the N-terminal region of the fusion protein encoded by
 CC the multicloning site of the poliovirus vector PMOV1.3. The sequence C
 CC spans the unique initiation site, the multicloning site, the protease C
 CC cleavage site and part of the N-terminus of the fusion polypeptide.
 CC The replication-competent recombinant viruses are useful as vaccines
 CC against bacterial, viral, fungal, and yeast infections, parasitic
 CC diseases, cancer and allergies.

XX Sequence 31 AA;

Query Match 100.0%; Score 30; DB 20; Length 31;
 Best Local Similarity 100.0%; Pred. No. 6.3;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 22 EALFOG 27

RESULT 8

AAV44189
 ID AAV44189 standard; peptide; 31 AA.

AC AAV44189;

DT 01-FEB-2000 (first entry)

DE N-terminus of fusion protein from modified poliovirus vector pMOV2.5.
 KM Antimicrobial; cytostatic; antiallergic; replication; recombinant;
 KM poliovirus; vector; expression; exogenous; protease; cleavage site;
 KM genome; primer; PCR; amplification; vaccine; bacterium; fungus; yeast;
 KM infection; parasitic disease; cancer; allergy.

OS Poliovirus.
 OS Synthetic.

PN USS965124-A.

PD 12-OCT-1999.

PF 31-JAN-1995; 9505-0381637.

PR 08-DEC-1992; 9205-0986729.

PR 06-DEC-1991; 9105-0804993.

PR 18-SEP-1992; 9205-0947790.

PA (WHEC) WHITEHEAD INST BIOMEDICAL RES.
 PA (AMCY) AMERICAN CYANAMID CO.

PI Feinberg M, Weeks-Levy CL, Rellily PA, Andlino R;

DR WPI: 1999-632429/54.

DR N-PSDB; AA220657.

PT Replication-competent recombinant poliovirus for use as vaccine against
 PT bacterial, viral, fungal, and yeast infections, parasitic diseases,
 PT cancer and allergies.

PS Example 2; Fig 2C; 40pp; English.

XX The invention relates to a replication-competent recombinant poliovirus
 CC vector which is able to express an exogenous polypeptide as a component
 CC of a recombinant polypeptide precursor. The precursor molecule also
 CC contains a protease cleavage allowing the precursor to be subsequently
 CC proteolytically processed by the polio protease 3C to release the
 CC exogenous protein. Preferably the exogenous sequence and the cleavage
 CC site sequence are inserted into the vector sequence at a location in
 CC the genome of the parent poliovirus such that they do not disrupt a
 CC polioviral sequence necessary for polioviral replication. This sequence
 CC corresponds to the N-terminal region of the fusion protein encoded by
 CC the multicloning site of the poliovirus vector pMOV2.5. The sequence
 CC spans the unique initiation site, the multicloning site, the protease C
 CC cleavage site and part of the N-terminus of the fusion polypeptide.
 CC The replication-competent recombinant viruses are useful as vaccines
 CC against bacterial, viral, fungal, and yeast infections, parasitic
 CC diseases, cancer and allergies.

SQ Sequence 31 AA;

Query Match 100.0%; Score 30; DB 20; Length 31;
 Best Local Similarity 100.0%; Pred. No. 6.3;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 22 EALFOG 27

RESULT 9
 ABB92703
 ID ABB92703 standard; Protein; 989 AA.

AC ABB92703;

DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 1914.

KM Herbicidal; plant; agriculture; herbicide.

OS Arabidopsis thaliana.

PN MO200210210-A2.

PD 07-FEB-2002.

PF 28-AUG-2001; 2001WO-EP09892.

PR 28-AUG-2001; 2001WO-EP09892.

PA (FARB) BAYER AG.

PI Tietjen K, Weidner M;

DR WPI: 2002-269010/31.

PT Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms.

PS Claim 5; SEQ ID NO 1914; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.

SQ Sequence 989 AA;

Query Match 100.0%; Score 30; DB 23; Length 989;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 235 EALFOG 240

RESULT 10
 AAR12141
 ID AAR12141 standard; Protein; 2185 AA.

AC AAR12141;

DT 05-AUG-1991 (first entry)

DE Enteroviral polypeptide.

KM Enteroviruses; monoclonal antibodies; myocarditis; myositis;
 KM meningitis; encephalitis; pancreatitis; post viral fatigue.

OS Enterovirus sp.


```

PN DE3939200-A.
XX
XX 29-MAY-1991.
PD
XX 27-NOV-1989; 89DE-3939200.
PF
XX 27-NOV-1989; 89DE-3939200.
PR
XX (PLAC ) MAX PLANCK GES WISSENSCH.
PA
XX Kandolf R:
PI
XX WPI: 1991-165150/23.
DR
XX N-PSDB; AA011816.
XX
XX New enteroviral polypeptide for raising group specific antibodies
PT - for detecting any type of enterovirus in blood or serum, and
PT new DNA encoding it
XX
XX Claim 1; pages 14-15; 26pp; German.
XX
XX This enteroviral polypeptide is used to raise poly- or monoclonal
CC antibodies (Abs). These are useful in assays for detecting entero-
CC virus specific antigens, as an indication of enteroviral disease.
CC All 70 serotypes of the enteroviral family can be detected.
CC Diseases such as myocarditis, myositis, meningitis, encephalitis
CC and pancreatitis can be diagnosed using the Abs.
XX
XX Sequence 2185 AA;
SQ
Query Match 100.0%; Score 30; DB 12; Length 2185;
Best Local Similarity 100.0%; Pred. NO. 4.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 EALFOG 6
DB 1425 EALFOG 1430
IIIIII
RESULT 11
AAR22210
ID AAR22210 standard; Protein; 2206 AA.
XX
XX AAR22210;
AC
XX 14-JUL-1992 (first entry)
DT
XX True type 3 poliovirus protein from LED3.
DE
XX RNA virus; error reduction.
XX
XX Poliovirus.
OS
XX MO9203538-A.
PN
XX 05-MAR-1992.
PD
XX 20-AUG-1991; 91WO-US05890.
PF
XX 20-AUG-1990; 90US-0570000.
PR
XX 20-AUG-1990; 90US-0569916.
XX
XX (UYNV-) COLUMBIA UNIV NEW Y.
PA
XX Racanietello V, Tatem JM, Weekeslevy CL;
PI
XX WPI: 1992-096882/12.
DR
XX N-PSDB; AA022965.
XX
XX New vaccine against infectious polio-virus comprises RNA virus -
PT for producing RNA virus cDNA and viable RNA virus
PT
XX Disclosure; fig 6; 110pp; English.
PS

```

```

XX
XX The protein sequence was deduced from the cDNA sequence of p3
CC poliovirus obt'd as in AA022965. The cDNA sequence is that of a
CC true RNA virus, i.e. the cDNA directs the produ. of a viable
CC RNA virus which is phenotypically similar to the source virus.
CC The full length cDNA in pLED3 was infectious. In vitro
CC transcription of pLED3 cDNA using T7 RNA polymerase produced
CC RNAs which possessed several erroneous amino acids. The RNA
CC viruses are used in vaccines against polio. The screening method
CC can be used during amplification of the source virus for vaccine
CC produ. to ensure maintenance of C at position 2493 in the viral
CC genome i.e. increasing the attenuation. The new prod. overcomes
CC the problem of errors introduced during replication of ss RNA,
CC which is much higher than for ds DNA.
XX
XX Sequence 2206 AA;
SQ
Query Match 100.0%; Score 30; DB 13; Length 2206;
Best Local Similarity 100.0%; Pred. NO. 4.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 EALFOG 6
DB 1449 EALFOG 1454
IIIIII
RESULT 12
AAP20037
ID AAP20037 standard; Protein; 2209 AA.
XX
XX AAP20037;
AC
XX 05-AUG-1992 (first entry)
DT
XX Sequence encoded by a full-length cDNA copy of the poliovirus genome
DE in plasmid pVR106.
XX
XX Poliovirus; picornavirus; vaccine; antigen; immunogen.
XX
XX Human poliovirus.
OS
XX
XX Key Location/Qualifiers
FH 1..69
FT /label= P4
FT 70..341
FT /label= VP2
FT 342..579
FT /label= VP3
FT 580..881
FT /label= VP1
FT 882..1030
FT /label= 3b
FT 1031..1127
FT /label= 5b
FT 1128..1456
FT /label= X
FT 1457..1543
FT /label= 1b
FT 1544..1565
FT /label= VPg
FT 1566..1748
FT /label= 2
FT 1749..2209
FT /label= 4(p(63))
XX
XX MO8203632-A.
PN
XX 28-OCT-1982.
PD
XX 12-NOV-1981; 81WO-3200525.
PF
XX 12-NOV-1981; 81US-0320525.
PR 20-APR-1981; 81US-0255879.
XX

```

XX (MASI) MASSACHUSETTS INST TECH.
 PA
 XX
 PI Baltimore D, Racanietello VR;
 XX
 DR WPI; 1982-95059E/44 (95059E)..
 DR N-PSDB: AAN20042.
 XX
 PT Prodn. of cDNA representing viral RNA sequences - by
 PT transcription, insertion into vector and host cell transformation
 XX
 PS Example; Table 1, pages 25-31; 50pp; English.
 XX
 CC Plasmid pVR106 was produced by combining plasmids pVR104 and pVR105.
 CC It contains a full-length cDNA copy of the poliovirus genome. E.coli
 CC HB101 contg. this plasmid has been registered as ATCC 31844. The
 CC full-length poliovirus cDNA molecule is itself infectious and can be
 CC introduced into cells and these cultured to produce RNA virus.
 CC Alternatively, the infectious cDNA can be treated with mutagens and
 CC the altered material used to infect cells so that attenuated viral
 CC RNA is prod. and this used to make vaccines. For antibody prod.,
 CC cDNA capable of directing antigen prodn. is selected and isolated and
 CC incorporated into cells which are incubated to produce RNA antigen.
 XX
 SQ Sequence 2209 AA;
 OY 1 EALFQG 6
 Db 1452 EALFQG 1457
 Query Match 100.0%; Score 30; DB 3; Length 2209;
 Best Local Similarity 100.0%; Pred. No. 4.8e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 13
 ID AAY82677 standard; Protein: 168 AA.
 AC AAY82677;
 XX
 DT 09-AUG-2000 (first entry)
 XX
 DE Sheep erythropoietin protein sequence.
 XX
 KW Erythropoietin; EPO; G-CSF; granulocyte colony stimulating factor;
 KW wobbler; codon altered gene; shuffling; modification; vaccine;
 KW insulin; peptide hormone; growth factor; cytokine; interferon;
 KW interleukin; leukaemia inhibitory factor; oncostatin M;
 KW transcription activator; expression activator; infectious organism.
 XX
 OS Ovis sp.
 OS
 PN WO200018906-A2.
 XX
 PD 06-APR-2000.
 XX
 PF 28-SEP-1999; 99WO-US22588.
 XX
 PR 29-SEP-1998; 98US-0102362.
 PR 29-JAN-1999; 99US-0117729.
 PR 05-FEB-1999; 99US-0118813.
 PR 24-JUN-1999; 99US-0141049.
 XX
 PA (MAXY-) MAXYGEN INC.
 PI Patten PA, Liu L, Stemmer WPC;
 XX
 DR WPI; 2000-303449/26.
 XX
 PT Novel methods for recombining codon-altered libraries of nucleic acids
 PT used to produce new proteins and new vectors with reduced rates of
 PT reversion to wild type -

XX Example; Fig 3; 92pp; English.
 PS
 XX
 CC A method has been developed of making codon altered nucleic acids (Mas),
 CC comprising providing a NA sequence (NA1) which encodes a polypeptide
 CC (P1), providing codon altered NA sequences, each encoding P1 or a
 CC modified form of it, and recombining the codon altered NA sequences to
 CC produce a target codon altered NA which encodes a second protein.
 CC The method of the invention can be used for recombining codon-altered
 CC libraries of nucleic acids to produce new proteins, which have
 CC improvements in a desirable characteristic. Target nucleic acids
 CC include those coding for therapeutic proteins such as erythropoietin
 CC (EPO), insulin, peptide hormones, growth factors, cytokines, interferons,
 CC interleukins, leukaemia inhibitory factor, and oncostatin M, as well as
 CC transcription and expression activators and proteins from infectious
 CC organisms for use as vaccines. The method can also be used to produce
 CC attenuated viruses which have reduced rates of reversion to wild type.
 CC The present sequence represents an EPO protein sequence, which is used
 CC in an example from the present invention.
 XX
 SQ Sequence 168 AA;
 OY 1 EALFQG 6
 Db 72 EALFQG 77
 Query Match 93.3%; Score 28; DB 21; Length 168;
 Best Local Similarity 83.3%; Pred. No. 1e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

RESULT 14
 ID AAP80131 standard; Protein: 2164 AA.
 AC AAP80131;
 XX
 DT 18-OCT-1990 (first entry)
 XX
 DE Peptides translated from Human Rhinovirus Strain 89 (HRV89) ORF.
 XX
 KW human rhinovirus Strain 89 (HRV89); immune system stimulation; ss.
 XX
 OS synthetic.
 OS
 FH Key
 FH Location/Qualifiers
 FT Peptide 1..69
 FT /label-VP4
 FT 70..336
 FT /label-VP2
 FT 337..574
 FT /label-VP3
 FT 575..872
 FT /label-VP1
 FT 873..1008
 FT /label-P2-A
 FT 1009..1103
 FT /label-P2-B
 FT 1104..1500
 FT /label-P2-C
 FT 1501..1521
 FT /label-VPg
 FT 1522..1704
 FT /label-protease
 FT 1705..3784
 FT /label-polymerase
 FT Peptide
 XX
 PN DE3628658-A.
 XX
 PD 03-MAR-1988.
 XX
 PF 23-AUG-1986; 86DE-3628658.
 XX

```

PR 23-AUG-1986; 86DE-3628658.
XX
XX (BOEH ) BOEHRINGER INGELHEIM.
XX
XX Duchler M, Skern T, Sommergruber W, Neubauer C, Grundler P, Blaas D;
PI Kuchler E;
XX
XX WPI: 1988-064926/10.
DR N-PSDB: AAN80153.
XX
XX New DNA coding for proteins of human rhinovirus 89 -
PT and new polypeptides for stimulating immunity of blocking cell
PT receptors
XX
XX Disclosure: ; P; German.
XX
XX The viral genome is RNA with a single large open reading frame encoding
CC 10 viral peptides. These peptides can be used therapeutically to
CC stimulate the immune system in response to HRV89 and to bind to/block
CC cellular receptors.
XX
XX SQ Sequence 2164 AA;
XX
XX Query Match 93.3%; Score 28; DB 9; Length 2164;
XX Best Local Similarity 83.3%; Pred. No. 1.4e+03;
XX Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 EALFOG 6
XX
XX DB 1420 EALFOG 1425
XX
XX RESULT 15
XX AAP81045
XX ID AAP81045 standard; Protein: 2164 AA.
XX
XX AC AAP81045;
XX
XX DT 21-AUG-1991 (first entry)
XX
XX DE Sequence of the viral proteins VP1-VP4, P2A-P2C, PCA-P3C encoded by
XX the genomic RNA of rhinovirus strain HRV89.
XX
XX KM Vaccine; diagnosis.
XX
XX OS Rhinovirus.
XX
XX Key Location/Qualifiers
XX FH 1.69
XX FT /label- VP4
XX FT 70..337
XX FT /label- VP2
XX FT 338..575
XX FT /label- VP3
XX FT 576..872
XX FT /label- VP1
XX FT 873..1007
XX FT /label- P2-A
XX FT 1008..1101
XX FT /label- P2-B
XX FT 1102..1421
XX FT /label- PC-2
XX FT 1422..1496
XX FT /label- P3-A
XX FT 1497..1517
XX FT /label- VPg
XX FT 1518..1678
XX FT /label- PROTEASE
XX FT 1679..2163
XX FT /label- POLYMERASE
XX
XX EP261403-A.
XX
XX

```

```

PD 30-MAR-1988.
XX
XX PF 20-AUG-1987; 87EP-0112104.
XX
XX PF 17-JAN-1987; 87DE-3701301.
XX
XX PR 23-AUG-1986; 86DE-3628658.
XX
XX PA (BOEH ) BOEHRINGER INGELHEIM.
XX
XX PI Duechler M, Skern T, Sommergruber W, Neubauer C, Grundler P, Blaas;
PI D, Kuchler E, Frascl L, Zorn M;
XX
XX WPI: 1988-085735/13.
DR N-PSDB: AAN81390.
XX
XX New DNA corresponding to viral RNA of rhino-virus HRV89 - useful for
PT prodn. of polypeptide(s) for stimulating immune system against HRV
PT 89.
XX
XX PS Clatm 8; Flg 4; 66pp; German.
XX
XX CC The viral proteins are used for stimulating a protective immune
XX response and for blocking cellular receptors. Abs against them
CC are useful for assay and purificn. of the corresp. antigen, and
CC can also be used for the therapeutic and diagnostic applications.
XX
XX SQ Sequence 2164 AA;
XX
XX Query Match 93.3%; Score 28; DB 9; Length 2164;
XX Best Local Similarity 83.3%; Pred. No. 1.4e+03;
XX Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 EALFOG 6
XX
XX DB 1420 EALFOG 1425
XX

```

Search completed: May 20, 2003, 11:33:56
Job time : 1.50658 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:30:54 ; Search time 0.19373 Seconds
(without alignments)
912.939 Million cell updates/sec

Title: US-10-081-408-6

Perfect score: 30

Sequence: 1 EALFQG 6

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched:

262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters:

262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries:

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCRTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/beckfilesl.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	30	100.0	11	2	US-08-381-637-27	Sequence 27, Appl
2	30	100.0	26	2	US-08-381-637-37	Sequence 29, Appl
3	30	100.0	31	2	US-08-381-637-29	Sequence 35, Appl
4	30	100.0	31	2	US-08-381-637-35	Sequence 2, Appl
5	30	100.0	2206	1	US-07-852-260-2	Sequence 2, Appl
6	30	100.0	2206	2	US-08-461-503-2	Sequence 2, Appl
7	30	100.0	2206	4	US-08-465-250-2	Sequence 2, Appl
8	27	90.0	323	6	5185254-4	Patent No. 5185254
9	27	90.0	365	3	US-09-113-309-2	Sequence 2, Appl
10	27	90.0	365	4	US-09-521-109-2	Sequence 2, Appl
11	27	90.0	365	4	US-09-562-332-2	Sequence 2, Appl
12	26	86.7	8	4	US-09-548-372D-62	Sequence 62, Appl
13	26	86.7	8	4	US-09-548-372D-62	Sequence 62, Appl
14	26	86.7	32	1	US-08-190-802A-185	Sequence 185, App
15	26	86.7	32	4	US-08-477-346-185	Sequence 185, App
16	26	86.7	32	4	US-08-473-089-185	Sequence 185, App
17	26	86.7	32	4	US-08-487-072A-185	Sequence 185, App
18	26	86.7	113	3	US-09-023-221A-20	Sequence 20, Appl
19	26	86.7	113	4	US-09-282-352A-20	Sequence 20, Appl
20	26	86.7	126	4	US-09-240-374-152	Sequence 152, App
21	26	86.7	127	4	US-09-240-374-152	Sequence 152, App
22	26	86.7	353	3	US-08-986-485-6	Sequence 139, App
23	26	86.7	353	3	US-08-986-485-6	Sequence 6, Appl
24	26	86.7	603	1	US-08-190-802A-50	Sequence 50, Appl
25	26	86.7	603	4	US-08-473-346-50	Sequence 50, Appl
26	26	86.7	603	4	US-08-473-089-50	Sequence 50, Appl
27	26	86.7	1134	3	US-08-487-072A-50	Sequence 50, Appl
					US-08-726-214-2	Sequence 2, Appl

28	26	86.7	1568	4	US-09-181-706-2	Sequence 2, Appl
29	26	86.7	1568	4	US-09-458-791-2	Sequence 2, Appl
30	26	86.7	1568	4	US-09-459-066-2	Sequence 2, Appl
31	25	83.3	19	1	US-08-383-753-63	Sequence 63, Appl
32	25	83.3	19	2	US-08-586-772-63	Sequence 63, Appl
33	25	83.3	19	2	US-08-959-512-63	Sequence 63, Appl
34	25	83.3	19	4	US-09-512-983-63	Sequence 63, Appl
35	25	83.3	21	1	US-08-828-323-21	Sequence 21, Appl
36	25	83.3	21	4	US-08-828-323-21	Sequence 21, Appl
37	25	83.3	178	4	US-09-315-689-5	Sequence 5, Appl
38	25	83.3	182	4	US-09-561-500-14	Sequence 14, Appl
39	25	83.3	182	4	US-09-561-108-14	Sequence 14, Appl
40	25	83.3	182	4	US-09-315-689-3	Sequence 3, Appl
41	25	83.3	182	4	US-09-561-526-14	Sequence 14, Appl
42	25	83.3	183	4	US-09-206-059-2	Sequence 2, Appl
43	25	83.3	184	3	US-08-741-411-1	Sequence 1, Appl
44	25	83.3	260	3	US-09-025-059-3	Sequence 3, Appl
45	25	83.3	260	3	US-09-070-526-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-381-637-27
Sequence 27, Application US/08381637
Patent No. 5965124
GENERAL INFORMATION:
APPLICANT: Mark Feinberg, Paul Andino, Carolyn Louise
TITLE OF INVENTION: Recombinant Vaccines and Method of
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESSES:
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: 2 Millita Drive
CITY: Lexington
STATE: MA
COUNTRY: US
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/381,637
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/986,729
FILING DATE: 08-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WH191-01AA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 11 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-381-637-27
Query Match 100.0%; Score 30; DB 2; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EALFQG 6
|||||

Db 4 EALFOG 9

RESULT 2

US-08-381-637-37
Sequence 37, Application US/08381637

Patent No. 5965124

GENERAL INFORMATION:

APPLICANT: Mark Feinberg, Raul Andino, Carolyn Louise

TITLE OF INVENTION: Recombinant Vaccines and Method of

NUMBER OF SEQUENCES: 39

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: 2 Militia Drive

CITY: Lexington

STATE: MA

COUNTRY: US

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/381,637

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/986,729

FILING DATE: 08-DEC-1992

ATTORNEY/AGENT INFORMATION:

NAME: Granahan, Patricia

REGISTRATION NUMBER: 32,227

REFERENCE/DOCKET NUMBER: WHI91-01AA

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 861-6240

TELEFAX: (617) 861-9540

INFORMATION FOR SEQ ID NO: 37:

SEQUENCE CHARACTERISTICS:

LENGTH: 26 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-381-637-37

Query Match 100.0%; Score 30; DB 2; Length 26;

Best Local Similarity 100.0%; Pred. No. 1.1;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6

Db 1 EALFOG 6

RESULT 3

US-08-381-637-29
Sequence 29, Application US/08381637

Patent No. 5965124

GENERAL INFORMATION:

APPLICANT: Mark Feinberg, Raul Andino, Carolyn Louise

TITLE OF INVENTION: Recombinant Vaccines and Method of

NUMBER OF SEQUENCES: 39

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: 2 Militia Drive

CITY: Lexington

STATE: MA

COUNTRY: US

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/381,637

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/986,729

FILING DATE: 08-DEC-1992

ATTORNEY/AGENT INFORMATION:

NAME: Granahan, Patricia

REGISTRATION NUMBER: 32,227

REFERENCE/DOCKET NUMBER: WHI91-01AA

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 861-6240

TELEFAX: (617) 861-9540

INFORMATION FOR SEQ ID NO: 29:

SEQUENCE CHARACTERISTICS:

LENGTH: 31 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-381-637-29

Query Match 100.0%; Score 30; DB 2; Length 31;

Best Local Similarity 100.0%; Pred. No. 1.4;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6

Db 22 EALFOG 27

RESULT 4

US-08-381-637-35
Sequence 35, Application US/08381637

Patent No. 5965124

GENERAL INFORMATION:

APPLICANT: Mark Feinberg, Raul Andino, Carolyn Louise

TITLE OF INVENTION: Recombinant Vaccines and Method of

NUMBER OF SEQUENCES: 39

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: 2 Militia Drive

CITY: Lexington

STATE: MA

COUNTRY: US

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/381,637

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/986,729

FILING DATE: 08-DEC-1992

ATTORNEY/AGENT INFORMATION:

NAME: Granahan, Patricia

REGISTRATION NUMBER: 32,227

REFERENCE/DOCKET NUMBER: WHI91-01AA

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 861-6240

TELEFAX: (617) 861-9540

INFORMATION FOR SEQ ID NO: 35:

SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-381-637-35

Query Match 100.0%; Score 30; DB 2; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
DB 22 EALFOG 27

RESULT 5
US-07-852-260-2
Sequence 2, Application US/07852260
Patent No. 5525715
GENERAL INFORMATION:
APPLICANT: Racanelli, Vincent
APPLICANT: Tatem, Joanne M.
APPLICANT: Weeks-Levy, Carolyn L.
TITLE OF INVENTION: METHODS FOR PRODUCING RNA VIRUSES FROM
TITLE OF INVENTION: CDNA
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/852,260
FILING DATE: 19920619
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 36607-B-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2206 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-852-260-2

Query Match 100.0%; Score 30; DB 1; Length 2206;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
DB 1449 EALFOG 1454

RESULT 6
US-08-461-503-2
Sequence 2, Application US/08461503
Patent No. 5834302
GENERAL INFORMATION:

APPLICANT: Racanelli, Vincent
APPLICANT: Tatem, Joanne M.
APPLICANT: Weeks-Levy, Carolyn L.
TITLE OF INVENTION: METHODS FOR PRODUCING RNA VIRUSES
TITLE OF INVENTION: FROM CDNA
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,503
FILING DATE: 5-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 36607-D-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2206 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-461-503-2

Query Match 100.0%; Score 30; DB 2; Length 2206;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
DB 1449 EALFOG 1454

RESULT 7
US-08-465-250-2
Sequence 2, Application US/08465250
Patent No. 6136570
GENERAL INFORMATION:
APPLICANT: Racanelli, Vincent
APPLICANT: Tatem, Joanne M.
APPLICANT: Weeks-Levy, Carolyn L.
TITLE OF INVENTION: METHODS FOR PRODUCING RNA VIRUSES FROM
TITLE OF INVENTION: CDNA
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release 1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,250
FILING DATE: 6-JUN-1995

```
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: White, John P.
/ REGISTRATION NUMBER: 28,678
/ REFERENCE/DOCKET NUMBER: 36607-E-PCT-US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 278-0400
/ TELEFAX: (212) 391-0525
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2206 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-08-465-250-2

Query Match
Best Local Similarity 100.0%; Score 30; DB 4; Length 2206;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
Db 1449 EALFOG 1454

RESULT 8
5185254-4
/ APPLICANT: LINNENBACH, ALBAN
/ TITLE OF INVENTION: GENE FAMILY OF TUMOR-ASSOCIATED ANTIGENS
/ NUMBER OF SEQUENCES: 5
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/07/291,583
/ FILING DATE: 29-DEC-1988
/ SEQ ID NO:4
/ LENGTH: 323
5185254-4

Query Match
Best Local Similarity 90.0%; Score 27; DB 6; Length 323;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
Db 233 EALFOG 238

RESULT 9
US-09-113-309-2
/ Sequence 2, Application US/09113309A
/ Patent No. 6110738
/ GENERAL INFORMATION:
/ APPLICANT: Zhou, Shihlin
/ APPLICANT: Zavel, Leigh
/ APPLICANT: Vogelstein, Bert
/ APPLICANT: Kinzler, Kenneth
/ TITLE OF INVENTION: Human Fast-1 Gene
/ FILE REFERENCE: 01107.10898
/ CURRENT APPLICATION NUMBER: US/09/113,309A
/ CURRENT FILING DATE: 1998-07-10
/ NUMBER OF SEQ ID NOS: 19
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 2
/ LENGTH: 365
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-113-309-2

Query Match
Best Local Similarity 90.0%; Score 27; DB 3; Length 365;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 EALFOG 6
Db 328 DALFOG 333

RESULT 10
US-09-521-109-2
/ Sequence 2, Application US/09521109
/ Patent No. 6225441
/ GENERAL INFORMATION:
/ APPLICANT: Zhou, Shihlin
/ APPLICANT: Zavel, Leigh
/ APPLICANT: Vogelstein, Bert
/ APPLICANT: Kinzler, Kenneth
/ TITLE OF INVENTION: Human Fast-1 Gene
/ FILE REFERENCE: 01107.10898
/ CURRENT APPLICATION NUMBER: US/09/521,109
/ CURRENT FILING DATE: 2000-03-07
/ PRIOR APPLICATION NUMBER: 09/113,309
/ PRIOR FILING DATE: 1998-07-10
/ NUMBER OF SEQ ID NOS: 19
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 2
/ LENGTH: 365
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-521-109-2

Query Match
Best Local Similarity 90.0%; Score 27; DB 4; Length 365;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
Db 328 DALFOG 333

RESULT 11
US-09-562-332-2
/ Sequence 2, Application US/09562332
/ Patent No. 6372434
/ GENERAL INFORMATION:
/ APPLICANT: Zhou, Shihlin
/ APPLICANT: Zavel, Leigh
/ APPLICANT: Vogelstein, Bert
/ APPLICANT: Kinzler, Kenneth
/ TITLE OF INVENTION: Human Fast-1 Gene
/ FILE REFERENCE: 01107.10898
/ CURRENT APPLICATION NUMBER: US/09/562,332
/ CURRENT FILING DATE: 2000-05-01
/ PRIOR APPLICATION NUMBER: 09/113,309
/ PRIOR FILING DATE: 1998-07-10
/ NUMBER OF SEQ ID NOS: 19
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 2
/ LENGTH: 365
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-562-332-2

Query Match
Best Local Similarity 90.0%; Score 27; DB 4; Length 365;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
Db 328 DALFOG 333

RESULT 12
US-09-548-372D-62
/ Sequence 62, Application US/09548372D
/ Patent No. 6420534
```


GENERAL INFORMATION:
; APPLICANT: GURNEY ET AL.
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR AND USES
; FILE REFERENCE: 29915/62801
; CURRENT APPLICATION NUMBER: US/09/548,372D
; CURRENT FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: US 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 62
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide

US-09-548-372D-62

Query Match 86.7%; Score 26; DB 4; Length 8;
Best Local Similarity 83.3%; Pred. No. 1.9e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EALFOG 6
1 |||||
DB 2 EYLFOG 7

RESULT 13

US-09-548-367D-62

Sequence 62, Application US/09548367D
Patent No. 6440698
GENERAL INFORMATION:
; APPLICANT: GURNEY ET AL.
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR AND USES
; FILE REFERENCE: 29915/6280H
; CURRENT APPLICATION NUMBER: US/09/548,367D
; CURRENT FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: US 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 62
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide

US-09-548-367D-62

Query Match 86.7%; Score 26; DB 4; Length 8;
Best Local Similarity 83.3%; Pred. No. 1.9e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EALFOG 6
1 |||||
DB 2 EYLFOG 7

RESULT 14

US-08-190-802A-185

Sequence 185, Application US/08190802A
Patent No. 5519003
GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: P.O. Box 60850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-0850
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/190,802A
; FILING DATE: 01-FEB-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 8600-0139
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 185:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Insulin-like growth factor bind.
; INDIVIDUAL ISOLATE: pro. complex-rat r1, Fig. 33

US-08-190-802A-185

Query Match 86.7%; Score 26; DB 1; Length 32;
Best Local Similarity 83.3%; Pred. No. 12;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EALFOG 6
1 |||||
DB 21 EGLFOG 26

RESULT 15

US-08-477-346-185

Sequence 185, Application US/08477346
Patent No. 6262023
GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,346
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,072
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2550-0025.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 32 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: Insulin-like growth factor bind.
INDIVIDUAL ISOLATE: pro. complex-rat IT, Fig. 33
US-08-477-346-185

Query Match 86.7%; Score 26; DB 4; Length 32;
Best Local Similarity 83.3%; Pred. No. 12;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EALFQG 6
| | | | |
Db 21 EQLFQG 26

Search completed: May 20, 2003, 11:40:58
Job time: 1.19337 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:37:59 ; Search time 0.955969 Seconds
(without alignments)

622.443 Million cell updates/sec

Title: US-10-081-408-6

Perfect score: 30

Sequence: 1.EALFGG 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep.*
5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	100.0	6	9 US-10-081-408-6	Sequence 5, Appli
2	30	100.0	9	9 US-10-081-408-20	Sequence 20, Appl
3	27	90.0	333	9 US-10-097-340-314	Sequence 314, App
4	27	90.0	533	10 US-09-864-761-48941	Sequence 48941, A
5	27	90.0	631	10 US-09-815-242-10711	Sequence 10711, A
6	27	90.0	707	9 US-09-884-696-21	Sequence 21, Appl
7	26	86.7	8	9 US-09-869-414-62	Sequence 62, Appl
8	26	86.7	8	10 US-09-794-927-62	Sequence 62, Appl
9	26	86.7	8	10 US-09-795-847-62	Sequence 62, Appl
10	26	86.7	8	10 US-09-794-743-62	Sequence 62, Appl
11	26	86.7	8	10 US-09-794-748-62	Sequence 62, Appl
12	26	86.7	8	10 US-09-794-925-62	Sequence 62, Appl
13	26	86.7	14	9 US-10-057-789-63	Sequence 62, Appl
14	26	86.7	14	9 US-10-057-789-64	Sequence 63, Appl
15	26	86.7	14	9 US-10-212-628-63	Sequence 64, Appl
16	26	86.7	14	9 US-10-212-628-64	Sequence 64, Appl
17	26	86.7	14	9 US-10-212-628-64	Sequence 64, Appl
18	26	86.7	20	9 US-10-057-789-61	Sequence 61, Appl
19	26	86.7	20	9 US-10-057-789-62	Sequence 62, Appl

20	26	86.7	20	9 US-10-212-628-61	Sequence 61, Appl
21	26	86.7	20	9 US-10-212-628-62	Sequence 62, Appl
22	26	86.7	113	10 US-09-738-599-70	Sequence 20, Appl
23	26	86.7	126	9 US-09-848-798-152	Sequence 152, App
24	26	86.7	127	9 US-09-848-798-159	Sequence 159, App
25	26	86.7	201	10 US-09-815-242-5217	Sequence 5217, Ap
26	26	86.7	202	10 US-09-764-864-1126	Sequence 1126, Ap
27	26	86.7	1024	10 US-09-864-921-97	Sequence 97, Appl
28	26	86.7	1024	10 US-09-841-739-2	Sequence 2, Appli
29	26	86.7	1204	10 US-09-841-739-5	Sequence 5, Appli
30	25	83.3	9	10 US-09-786-284-45	Sequence 21, Appl
31	25	83.3	9	10 US-09-796-294-45	Sequence 45, Appl
32	25	83.3	11	9 US-09-879-572A-23	Sequence 23, Appl
33	25	83.3	11	9 US-09-879-572A-25	Sequence 25, Appl
34	25	83.3	50	9 US-09-879-572A-32	Sequence 32, Appl
35	25	83.3	63	10 US-09-822-540A-1	Sequence 1, Appli
36	25	83.3	77	9 US-09-879-572A-30	Sequence 30, Appl
37	25	83.3	96	10 US-09-864-761-34961	Sequence 34961, A
38	25	83.3	132	9 US-09-764-891-4364	Sequence 4364, Ap
39	25	83.3	178	9 US-10-131-241-60	Sequence 60, Appl
40	25	83.3	179	9 US-10-131-241-57	Sequence 57, Appl
41	25	83.3	180	9 US-10-131-241-55	Sequence 56, Appl
42	25	83.3	181	9 US-10-131-241-55	Sequence 55, Appl
43	25	83.3	182	9 US-10-131-241-54	Sequence 54, Appl
44	25	83.3	182	10 US-09-998-831-14	Sequence 14, Appl
45	25	83.3	183	9 US-10-080-797-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-10-081-408-6
Sequence 6, Application US/10081408
Patent No. US20020160482A1
GENERAL INFORMATION:
APPLICANT: Abrahams, n , Lars
APPLICANT: Nilsson, Joakim
TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION
FILE REFERENCE: 13425-053001
CURRENT APPLICATION NUMBER: US/10/081,408
CURRENT FILING DATE: 2002-02-21
PRIOR FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: SE 0100625-3
PRIOR FILING DATE: 2001-02-23
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 20
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 6
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Protease cleavage site
US-10-081-408-6
Query Match 100.0%; Score 30; DB 9; Length 6;
Best Local Similarity 100.0%; Pred. No. 3.3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;
OY 1 EALFGG 6
Db 1 EALFGG 6
US-10-081-408-20
Sequence 20, Application US/10081408
Patent No. US20020160482A1
GENERAL INFORMATION:
APPLICANT: Abrahams, n , Lars
APPLICANT: Nilsson, Joakim
TITLE OF INVENTION: METHODS FOR PROTEIN PURIFICATION

FILE REFERENCE: 13425-053001
CURRENT APPLICATION NUMBER: US/10/081,408
CURRENT FILING DATE: 2002-02-21
PRIOR APPLICATION NUMBER: SE 0100625-3
PRIOR FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: US 60/272,247
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 20
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 20
LENGTH: 998
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant construct
US-10-081-408-20

Query Match 100.0%; Score 30; DB 9; Length 998;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
DB 259 EALFOG 264

RESULT 3
US-10-097-340-314
Sequence 314, Application US/10097340
Publication No. US20030087250A1
GENERAL INFORMATION:
APPLICANT: John MONAHAN
APPLICANT: Manjula GANNAVAPU
APPLICANT: Sebastian HOERSCH
APPLICANT: Shubhangi KAMATKAR
APPLICANT: Steve G. KOVATS
APPLICANT: Rachel E. MEYERS
APPLICANT: Michael MORRISSEY
APPLICANT: Peter OLANDT
APPLICANT: Ami SEN
APPLICANT: Peter VEIRY
APPLICANT: Gordon B. MILLS
APPLICANT: Robert C. BAST, JR.
APPLICANT: Karen LU
APPLICANT: Rosemarie SCHMANDT
APPLICANT: Xumel ZHAO
APPLICANT: Karen GLATT
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer
FILE REFERENCE: MRI-030
CURRENT APPLICATION NUMBER: US/10/097,340
CURRENT FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: 60/276,025
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/325,149
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/276,026
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/324,967
PRIOR FILING DATE: 2001/09/26
PRIOR APPLICATION NUMBER: 60/311,732
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: 60/325,102
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/323,580
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 363
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 314
LENGTH: 323
TYPE: PRT
ORGANISM: Homo sapiens
US-10-097-340-314

Query Match 90.0%; Score 27; DB 9; Length 323;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
DB 233 EALFOG 238

RESULT 4
US-09-864-761-48941
Sequence 48941, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FC
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 48941
LENGTH: 533
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC008174.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.63
OTHER INFORMATION: EST_HUMAN HIT: BF129490.1, EVALU 1.30e+00
OTHER INFORMATION: SWISSPROT HIT: P13666, EVALU 6.40e-02
US-09-864-761-48941

Query Match 90.0%; Score 27; DB 10; Length 533;
Best Local Similarity 83.3%; Pred. No. 2.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
1:||||
DB 494 EALFOG 499

RESULT 5
US-09-815-242-10711

; Sequence 10711, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trivick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10711
; LENGTH: 631
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10711

Query Match 90.0%; Score 27; DB 10; Length 631;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
1:||||
DB 89 EALFOG 94

RESULT 6
US-09-884-696-21

; Sequence 21, Application US/09884696
; Publication No. US20030035809A1
; GENERAL INFORMATION:
; APPLICANT: GEORGE, LITTLE W
; APPLICANT: ANGELIOS, JOHN A
; APPLICANT: HESS, JOHN F
; TITLE OF INVENTION: MORAXELLA BOVIS CYTOTOXIN, CYTOTOXIN GENE, ANTIBODIES
; TITLE OF INVENTION: AND VACCINES FOR PREVENTION AND TREATMENT OF MORAXELLA
; TITLE OF INVENTION: BOVIS INJECTIONS
; FILE REFERENCE: 481.06
; CURRENT APPLICATION NUMBER: US/09/884,696
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 21

; LENGTH: 707

; TYPE: PRT

; ORGANISM: Escherichia coli

Query Match 90.0%; Score 27; DB 9; Length 707;
Best Local Similarity 83.3%; Pred. No. 3.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
1:||||
DB 115 EALFOG 120

RESULT 7
US-09-869-414-62

; Sequence 62, Application US/09869414
; Publication No. US20030077226A1
; GENERAL INFORMATION:
; APPLICANT: Belinkowski et al.
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND I
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 28341/6280M
; CURRENT APPLICATION NUMBER: US/09/869,414
; CURRENT FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: 09/416,901
; PRIOR FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-869-414-62

Query Match 86.7%; Score 26; DB 9; Length 8;
Best Local Similarity 83.3%; Pred. No. 3.3e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EALFOG 6
1:||||
DB 2 EALFOG 7

RESULT 8
US-09-794-927-62

; Sequence 62, Application US/09794927
; Patent No. US20010016324A1
; GENERAL INFORMATION:
; APPLICANT: Gurney, Mark E.
; APPLICANT: Bienkowski, Michael J.
; APPLICANT: Helinikson, Robert L.
; APPLICANT: Parodi, Luis A.
; APPLICANT: Yan, Ridiang
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND
; TITLE OF INVENTION: USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 28341/6280FG
; CURRENT APPLICATION NUMBER: US/09/794,927
; CURRENT FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 09/416,901

```

; PRIOR FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 62
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-794-927-62
```

```
Query Match      86.7%; Score 26; DB 10; Length 8;
Best Local Similarity 83.3%; Pred. No. 3.3e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      1 EALFOG 6
        1 | 1111
Db       2 EVLFOG 7
```

```
RESULT 9
US-09-795-847-62
; Sequence 62, Application US/09795847
; Patent No. US20010018208A1
; GENERAL INFORMATION:
; APPLICANT: Gurney, Mark E.
; APPLICANT: Bienkowski, Michael J.
; APPLICANT: Heinrichson, Robert L.
; APPLICANT: Parodi, Luis A.
; APPLICANT: Yan, Riqiang
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND
; TITLE OF INVENTION: USES
; FILE REFERENCE: 28341/6280DE
; CURRENT APPLICATION NUMBER: US/09/795,847
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 09/416,901
; PRIOR FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 62
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-795-847-62
```

```
Query Match      86.7%; Score 26; DB 10; Length 8;
Best Local Similarity 83.3%; Pred. No. 3.3e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      1 EALFOG 6
        1 | 1111
Db       2 EVLFOG 7
```

```
RESULT 10
US-09-794-743-62
; Sequence 62, Application US/09794743
; Patent No. US20010021391A1
; GENERAL INFORMATION:
; APPLICANT: Gurney, Mark E.
; APPLICANT: Bienkowski, Michael J.
; APPLICANT: Heinrichson, Robert L.
; APPLICANT: Parodi, Luis A.
; APPLICANT: Yan, Riqiang
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND
; TITLE OF INVENTION: USES
; FILE REFERENCE: 28341/6280BC
; CURRENT APPLICATION NUMBER: US/09/794,743
; CURRENT FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 09/416,901
; PRIOR FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 62
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-794-743-62
```

```
Query Match      86.7%; Score 26; DB 10; Length 8;
Best Local Similarity 83.3%; Pred. No. 3.3e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      1 EALFOG 6
        1 | 1111
Db       2 EVLFOG 7
```

```
RESULT 11
US-09-794-748-62
; Sequence 62, Application US/09794748
; Patent No. US20020037315A1
; GENERAL INFORMATION:
; APPLICANT: Gurney, Mark E.
; APPLICANT: Bienkowski, Michael J.
; APPLICANT: Heinrichson, Robert L.
; APPLICANT: Parodi, Luis A.
; APPLICANT: Yan, Riqiang
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND
; TITLE OF INVENTION: USES
; FILE REFERENCE: 28341/6280UL
; CURRENT APPLICATION NUMBER: US/09/794,748
; CURRENT FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 09/416,901
; PRIOR FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: Patentln Ver. 2.0
```

SEQ ID NO 62
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-794-748-62

Query Match
Best Local Similarity 86.7%; Score 26; DB 10; Length 8;
Best Local Similarity 83.3%; Pred. No. 3.3e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EALFOG 6
1 ||||
DB 2 EVLFOG 7

RESULT 12
US-09-794-925-62
Sequence 62, Application US/09794925
Patent No. US20020064819A1
GENERAL INFORMATION:
APPLICANT: Gurney, Mark E.
APPLICANT: Bienkowski, Michael J.
APPLICANT: Heinrikson, Robert L.
APPLICANT: Parodi, Luis A.
APPLICANT: Yan, Riqiang
TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND USES
FILE REFERENCE: 28341/6280H1
CURRENT FILING DATE: 2001-02-27
CURRENT APPLICATION NUMBER: US/09/794,925
PRIOR FILING DATE: 1999-10-13
PRIOR APPLICATION NUMBER: 60/155,493
PRIOR FILING DATE: 1999-09-23
PRIOR APPLICATION NUMBER: 09/404,133
PRIOR FILING DATE: 1999-09-23
PRIOR APPLICATION NUMBER: PCT/US99/20881
PRIOR FILING DATE: 1999-09-23
PRIOR APPLICATION NUMBER: 60/101,594
PRIOR FILING DATE: 1998-09-24
NUMBER OF SEQ ID NOS: 73
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 62
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-794-925-62

Query Match
Best Local Similarity 86.7%; Score 26; DB 10; Length 8;
Best Local Similarity 83.3%; Pred. No. 3.3e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EALFOG 6
1 ||||
DB 2 EVLFOG 7

RESULT 13
US-09-681-442-62
Sequence 62, Application US/09681442
Patent No. US20020081634A1
GENERAL INFORMATION:
APPLICANT: Gurney, Mark E.
APPLICANT: Bienkowski, Michael J.
APPLICANT: Heinrikson, Robert L.
APPLICANT: Parodi, Luis A.
APPLICANT: Yan, Riqiang
TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND USES
TITLE OF INVENTION: THEREFOR

FILE REFERENCE: 28341/6280FG
CURRENT APPLICATION NUMBER: US/09/681,442
CURRENT FILING DATE: 2001-04-05
PRIOR APPLICATION NUMBER: 09/416,901
PRIOR FILING DATE: 1999-10-13
PRIOR APPLICATION NUMBER: 60/155,493
PRIOR FILING DATE: 1999-09-23
PRIOR APPLICATION NUMBER: 09/404,133
PRIOR FILING DATE: 1999-09-23
PRIOR APPLICATION NUMBER: PCT/US99/20881
PRIOR FILING DATE: 1999-09-23
PRIOR APPLICATION NUMBER: 60/101,594
PRIOR FILING DATE: 1998-09-24
NUMBER OF SEQ ID NOS: 73
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 62
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-681-442-62

Query Match
Best Local Similarity 86.7%; Score 26; DB 10; Length 8;
Best Local Similarity 83.3%; Pred. No. 3.3e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EALFOG 6
1 ||||
DB 2 EVLFOG 7

RESULT 14
US-10-057-789-63
Sequence 63, Application US/10057789
Publication No. US20030082522A1
GENERAL INFORMATION:
APPLICANT: Paul Haynes
APPLICANT: Jing Wei
APPLICANT: John Yates
APPLICANT: Nancy Andon
TITLE OF INVENTION: ANALYSIS OF COMPLEX PROTEIN MIXTURES
TITLE OF INVENTION: DIFFERENTIAL LABELING FOR QUANTITATIVE
FILE REFERENCE: NADIT.022A
CURRENT APPLICATION NUMBER: US/10/057,789
CURRENT FILING DATE: 2002-06-28
PRIOR APPLICATION NUMBER: US 60/264,576
PRIOR FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: US 60/305,232
PRIOR FILING DATE: 2001-07-13
NUMBER OF SEQ ID NOS: 311
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 63
LENGTH: 14
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthesized Peptide
US-10-057-789-63

Query Match
Best Local Similarity 86.7%; Score 26; DB 9; Length 14;
Best Local Similarity 83.3%; Pred. No. 8;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EALFOG 6
1 ||||
DB 7 EVLFOG 12

RESULT 15
US-10-057-789-64
Sequence 64, Application US/10057789
Publication No. US20030082522A1

GENERAL INFORMATION:
APPLICANT: Paul Haynes
APPLICANT: Jing Wei
APPLICANT: John Yates
APPLICANT: Nancy Andon
TITLE OF INVENTION: DIFFERENTIAL LABELING FOR QUANTITATIVE
FILE REFERENCE: NADII.022A
CURRENT APPLICATION NUMBER: US/10/057,789
CURRENT FILING DATE: 2002-06-28
PRIOR APPLICATION NUMBER: US 60/264,576
PRIOR FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: US 60/305,232
PRIOR FILING DATE: 2001-07-13
NUMBER OF SEQ ID NOS: 311
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 64
LENGTH: 14
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthesized Peptide
FEATURE:
NAME/KEY: VARIANT
LOCATION: 14
OTHER INFORMATION: Xaa - Ornithine
US-10-057-789-64

Query Match 86.7%; Score 26; DB 9; Length 14;
Best Local Similarity 83.3%; Pred. NO. 8;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EALFQG 6
Db 7 EVLFQG 12

Search completed: May 20, 2003, 12:02:22
Job time : 0.955969 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 11:09:54 ; Search time 0.231502 Seconds
(without alignments)
2491.578 Million cell updates/sec

Title: US-10-081-408-6

Perfect score: 30
Sequence: 1 EALFOG 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	100.0	989	2 T46183	zinc proteinase (E
2	30	100.0	1081	2 S15040	pleiotropic drug r
3	30	100.0	1374	1 GNNYEC	genome polypotein
4	30	100.0	2175	1 GNNYBE	genome polypotein
5	30	100.0	2182	1 GNNYB1	genome polypotein
6	30	100.0	2183	1 GNNYB4	genome polypotein
7	30	100.0	2185	1 GNNYB3	genome polypotein
8	30	100.0	2185	1 GNNYBT	genome polypotein
9	30	100.0	2185	1 JQ2021	genome polypotein
10	30	100.0	2185	1 GNNYSV	genome polypotein
11	30	100.0	2185	1 GNNYSV	genome polypotein
12	30	100.0	2193	2 S44194	genome polypotein
13	30	100.0	2193	2 S44251	polyprotein - echo
14	30	100.0	2193	2 S52919	polyprotein (1A, 1
15	30	100.0	2194	2 GNNYF7	genome polypotein
16	30	100.0	2201	1 GNNYF9	genome polypotein
17	30	100.0	2206	1 GNNY27	genome polypotein
18	30	100.0	2206	1 GNNY4P	genome polypotein
19	30	100.0	2206	2 S03822	genome polypotein
20	30	100.0	2207	1 GNNY1P	genome polypotein
21	30	100.0	2207	1 GNNY1P	genome polypotein
22	30	100.0	2207	1 GNNY1P	genome polypotein
23	30	100.0	2207	1 GNNY1P	genome polypotein
24	30	100.0	2209	1 GNNY2P	genome polypotein
25	30	100.0	2209	1 GNNY3P	genome polypotein
26	30	100.0	2214	1 A48548	genome polypotein
27	28	93.3	247	2 A87511	conserved hypotet
28	28	93.3	278	2 E83152	probable short-cha
29	28	93.3	425	2 T10702	phycocyan synthase

30	28	93.3	445	2 AG1556	Na ⁺ -transporting A
31	28	93.3	2164	1 GNNY89	genome polypotein
32	27	90.0	117	2 A95907	probable protein i
33	27	90.0	117	2 G95924	hypothetical prote
34	27	90.0	201	2 A82466	hypothetical prote
35	27	90.0	227	2 AB2008	hypothetical prote
36	27	90.0	256	2 T36793	probable oxidoredu
37	27	90.0	323	1 A48149	cytochrome-c oxida
38	27	90.0	338	2 A82209	carcinoma-associat
39	27	90.0	380	2 T40541	conserved hypotet
40	27	90.0	395	2 H70385	hypothetical prote
41	27	90.0	442	2 T18917	tryptophan-tRNA 11
42	27	90.0	442	2 T11804	hypothetical prote
43	27	90.0	452	2 S22199	translation initia
44	27	90.0	614	2 B43935	imidazoleglycerol-
45	27	90.0	614	2 B43935	spat protein - Bac

ALIGNMENTS

RESULT 1
T46183
zinc proteinase (EC 3.4.24.-) T8H10.70 - Arabidopsis thaliana
N:Alternate names: Insulysin homolog
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 17-Mar-2000
C:Accession: T46183
R:Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Mewes, H.W.; Lemcke, K.; Mayer
submitted to the Protein Sequence Database, January 2000
A:Reference number: 223014
A:Accession: T46183
A:Molecule type: DNA
A:Residues: 1-989 <BEN>
A:Cross-References: EMBL:AL133248; PIDN:CA66104.1
A:Experimental source: cultivar Columbia; BAC clone T8H10
C:Genetics:
A:Map position: 3
A:introns: 48/3; 74/1; 96/3; 127/2; 152/1; 166/3; 184/1; 225/1; 262/3; 317/1; 348/1
A:Note: T8H10.70
C:Superfamily: Insulysin
C:Keywords: hydrolase; metalloproteinase; zinc
F:1.75/Binding site: zinc (His) #status predicted
F:74/Active site: Glu #status predicted

Query Match 100.0%; Score 30; DB 2; Length 989;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
Db 235 EALFOG 240
RESULT 2
S15040
pleiotropic drug resistance control protein PR6 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein G3713; protein YGL016w
C:Species: Saccharomyces cerevisiae
C:Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 21-Jul-2000
R:Chen, W.; Balz, E.; Capleau, E.; Choder, M.; Goffeau, A.
Yeast 7, 287-299, 1991
A:Title: The DNA sequencing of the 17 kb HindIII fragment spanning the LEU1 and ARG
ntroiling pleiotropic drug resistance.
A:Reference number: S15040; MUID:91353083; PMID:1882553
A:Accession: S15040
A:Molecule type: DNA
A:Residues: 1-1081 <CH>
A:Cross-References: GB:557895; NID:9234319; PIDN:AA19613.1; PID:9234320
R:Hebling, U.; Hofmann, B.; Dellus, H.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64003

A:Accession: S64018
A:Molecule type: DNA
A:Residues: 1-1081 <HEB>
A:Cross-references: EMBL:272538; NID:91322476; PID:6243847; PID:91322477; MIPS:YGL016w
A:Experimental source: strain 5288C
C:Genetics:
A:Gene: SGD:PD86
A:Cross-references: SGD:S0002984; MIPS:YGL016w
A:Map position: 7L
C:Keywords: transmembrane protein
F:365-381/Domain: transmembrane #status predicted <TM2>
F:540-556/Domain: transmembrane #status predicted

Query Match 100.0%; Score 30; DB 2; Length 1081;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
|||||
DB 631 EALFOG 636

RESULT 3

GNNYFC

genome polypeptide - echovirus 11 (strain Gregory) (fragment)
N:Contains: carboxyl end of coat protein 1D; core protein 2A; core protein 2B; core prot
C:Species: echovirus 11
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jun-2000
C:Accession: A36642
R:Avunen, P.; Hyypia, T.
J. Gen. Virol. 71, 2133-2139, 1990
A:Title: Echoviruses include genetically distinct serotypes.
A:Reference number: A36642; MUID:91011360; PMID:2170575
A:Accession: A36642
A:Molecule type: genomic RNA
A:Residues: 1-1374 <ANV>
A:Cross-references: GB:D010582; GB:D01068; NID:9465394; PIDN:BA01439.1; PID:9465395
C:Superfamily: poliovirus genome polypeptide
C:Keywords: coat protein; core protein; genome-linked protein; nucleotidyltransferase; F
F:1-43/Product: coat protein 1D (fragment) #status predicted <VP1>
F:44-190/Product: core protein 2A #status predicted <2AP>
F:191-289/Product: core protein 2B #status predicted <2BP>
F:290-618/Product: core protein 2C #status predicted <2CP>
F:619-707/Product: core protein 3A #status predicted <P3A>
F:708-729/Product: genome-linked protein VPg #status predicted <VPg>
F:730-912/Product: proteinase #status predicted <PRTS>
F:913-1374/Product: RNA-directed RNA polymerase #status predicted <RPS>
F:710/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 30; DB 1; Length 1374;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
|||||
DB 614 EALFOG 619

RESULT 4

GNNYBE

genome polypeptide - bovine enterovirus (strain VG-5-27)
N:Contains: coat protein VP1; coat protein VP2; coat protein VP3; coat protein VP4; core
RNA-directed RNA polymerase (EC 2.7.7.48)
C:Species: bovine enterovirus
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jun-2000
C:Accession: A29824
R:Earle, J.A.P.; Skuce, R.A.; Fleming, C.S.; Hoey, E.M.; Martin, S.J.
J. Gen. Virol. 69, 253-263, 1988
A:Title: The complete nucleotide sequence of a bovine enterovirus.
A:Reference number: A29824; MUID:88117392; PMID:2828511
A:Accession: A29824
A:Molecule type: genomic RNA
A:Residues: 1-2175 <EAR>

A:Cross-references: GB:D00214; NID:92696866; PIDN:BA024003.1
C:Superfamily: poliovirus genome polypeptide
C:Keywords: coat protein; core protein; genome-linked protein; nucleotidyltransferase
F:1-69/Product: coat protein VP4 #status predicted <VP4>
F:70-317/Product: coat protein VP2 #status predicted <VP2>
F:318-558/Product: coat protein VP3 #status predicted <VP3>
F:560-840/Product: coat protein VP1 #status predicted <VP1>
F:841-990/Product: core protein P2-A #status predicted <P2A>
F:991-1089/Product: core protein P2-B #status predicted <P2B>
F:1090-1419/Product: core protein P2-C #status predicted <P2C>
F:1420-1508/Product: core protein P3-A #status predicted <P3A>
F:1509-1531/Product: genome-linked protein VPg #status predicted <VPg>
F:1532-1714/Product: proteinase #status predicted <PRTS>
F:1715-2175/Product: RNA-directed RNA polymerase #status predicted <RPS>
F:1511/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 30; DB 1; Length 2175;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
|||||
DB 1415 EALFOG 1420

RESULT 5

GNNYB1

genome polypeptide - coxsackievirus B1
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core
polymerase (EC 2.7.7.48)
C:Species: coxsackievirus B1
C:Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999
C:Accession: A26353
R:Itzuka, N.; Kuge, S.; Nomoto, A.
Virology 156, 64-73, 1987
A:Title: Complete nucleotide sequence of the genome of coxsackievirus B1.
A:Reference number: A26353; MUID:87122157; PMID:3027969
A:Accession: A26353
A:Molecule type: genomic RNA
A:Residues: 1-2182 <IIZ>
A:Cross-references: GB:M6560; NID:9323417; PIDN:AA00531.1; PID:9323418

C:Superfamily: poliovirus genome polypeptide
C:Keywords: coat protein; core protein; genome-linked protein; nucleotidyltransferase
F:1-69/Product: coat protein 1A #status predicted <P1A>
F:70-332/Product: coat protein 1B #status predicted <P1B>
F:333-570/Product: coat protein 1C #status predicted <P1C>
F:571-848/Product: coat protein 1D #status predicted <P1D>
F:849-998/Product: core protein 2A #status predicted <C2A>
F:999-1097/Product: core protein 2B #status predicted <C2B>
F:1098-1426/Product: core protein 2C #status predicted <C2C>
F:1427-1515/Product: protein 3A #status predicted <P3A>
F:1516-1537/Product: genome-linked protein VPg #status predicted <VPg>
F:1538-1720/Product: proteinase #status predicted <PRTS>
F:1721-2182/Product: RNA-directed RNA polymerase #status predicted <RNS>
F:1518/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 30; DB 1; Length 2182;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
|||||
DB 1422 EALFOG 1427

RESULT 6

GNNYB4

genome polypeptide - coxsackievirus B4
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core
polymerase (EC 2.7.7.48)
C:Species: coxsackievirus B4
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 16-Jul-1999
C:Accession: A27170; A43576; B45576

R.Jenkins, O.; Booth, J.D.; Minor, P.D.; Almond, J.W.
J. Gen. Virol. 68, 1835-1848, 1987
A:Title: The complete nucleotide sequence of coxsackievirus B4 and its comparison to other
A:Reference number: A27170; MUID:87253111; PMID:3037008
A:Accession: A27170
A:Molecule type: genomic RNA
A:Residues: 1-2183 <JEN>
R.Ramalingam, A.; Araki, H.; Bryant, S.; Hixson, A.
Virus Res. 23, 281-292, 1992
A:Title: Identification of candidate sequences that determine virulence in coxsackievirus
A:Reference number: A45576; MUID:92327853; PMID:1320798
A:Accession: A45576
A:Molecule type: genomic RNA
A:Residues: 569-696, 'T', 698-852 <RAM>
A:Cross-references: GB:S39291; NID:9250908; PIDN:AA82445.1; PID:9250909
A:Experimental source: B4; virulent strain
A:Note: sequence inconsistent with the nucleotide translation
A:Accession: B45576
A:Molecule type: genomic RNA
A:Residues: 70-203, 'A', 205-231, 'E', 233-269, 'A', 271-330 <RA2>
A:Cross-references: GB:S39291; NID:9250908; PIDN:AA82446.1; PID:9250910
A:Experimental source: B4; virulent strain
A:Note: sequence extracted from NCBI backbone (NCBIN:108105, NCBIP:108107)
C:Superfamily: poliovirus genome polypeptide
C:Keywords: coat protein; core protein; genome-linked protein; nucleotidyltransferase; F
F:1-65/Product: coat protein 1A #status predicted <C1A>
F:70-130/Product: coat protein 1B #status predicted <C1B>
F:131-568/Product: coat protein 1C #status predicted <C1C>
F:569-852/Product: coat protein 1D #status predicted <C1D>
F:853-999/Product: core protein 2A #status predicted <C2A>
F:1000-1098/Product: core protein 2B #status predicted <C2B>
F:1099-1427/Product: core protein 2C #status predicted <C2C>
F:1428-1516/Product: protein 3A #status predicted <P3A>
F:1517-1538/Product: genome-linked protein VPg #status predicted <VPG>
F:1539-1721/Product: proteinase #status predicted <PPS>
F:1722-2183/Product: RNA-directed RNA polymerase #status predicted <RPS>
F:1519/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 30; DB 1; Length 2183;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
Db 1423 EALFOG 1428

RESULT 7
GNNYB3
genome polypeptide - coxsackievirus B3 (strain Nancy)
N:Contains: coat protein VP1; coat protein VP2; coat protein VP3; coat protein VP4; core
EC 3.4.-.-; RNA-directed RNA polymerase (EC 2.7.7.48)
C:Species: coxsackievirus B3
A:Notes: host Homo sapiens (man)
C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 05-Jun-1998
A:Accession: A26354; B00725
R.Lindberg, A.M.; Stalhandske, P.O.K.; Petersson, U.
Virology 156, 50-63, 1987
A:Title: Genome of coxsackievirus B3
A:Reference number: A26354; MUID:87122156; PMID:3027968
A:Accession: A26354
A:Molecule type: genomic RNA
A:Residues: 1-2185 <LIN>
A:Cross-references: GB:K02709
R.Stalhandske, P.O.K.; Lindberg, M.; Petersson, U.
J. Virol. 51, 742-746, 1984
A:Title: Replicase gene of coxsackievirus B3.
A:Reference number: A00725; MUID:84292451; PMID:6088796
A:Accession: B00725
A:Molecule type: genomic RNA
A:Residues: 1724-1733, 'D', 1735-2185 <STA>
A:Cross-references: GB:M16572

C:Superfamily: poliovirus genome polypeptide
C:Keywords: coat protein; core protein; genome-linked protein; hydrolase; nucleotid
F:1-69/Product: coat protein VP4 #status predicted <VP4>
F:70-132/Product: coat protein VP2 #status predicted <VP2>
F:133-570/Product: coat protein VP3 #status predicted <VP3>
F:571-880/Product: coat protein VP1 #status predicted <VP1>
F:881-1001/Product: core protein P2-A #status predicted <P2A>
F:1002-1100/Product: core protein P2-B #status predicted <P2B>
F:1101-1429/Product: core protein P2-C #status predicted <P2C>
F:1430-1518/Product: core protein P3-A #status predicted <P3A>
F:1519-1540/Product: genome-linked protein VPg #status predicted <VPG>
F:1541-1723/Product: proteinase #status predicted <PPS>
F:1724-2185/Product: RNA-directed RNA polymerase #status predicted <RPS>
F:1521/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 30; DB 1; Length 2185;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
Db 1425 EALFOG 1430

RESULT 8
GNNYB3
genome polypeptide - coxsackievirus B3
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; cor
polymerase (EC 2.7.7.48)
C:Species: coxsackievirus B3
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jul-1999
R.Klump, W.M.; Bergmann, I.; Mueller, B.C.; Ameis, D.; Kandolf, R.
J. Virol. 64, 1573-1583, 1990
A:Title: Complete nucleotide sequence of infectious coxsackievirus B3 cDNA: two int
A:Reference number: A34664; MUID:90204667; PMID:2157045
A:Accession: A34664
A:Molecule type: genomic RNA
A:Residues: 1-2185 <KLU>
A:Cross-references: GB:M33854; NID:9323419; PIDN:AAA2931.1; PID:9323420
C:Superfamily: poliovirus genome polypeptide
C:Keywords: coat protein; core protein; genome-linked protein; nucleotidyltransfera
F:1-69/Product: coat protein 1A #status predicted <VP4>
F:70-132/Product: coat protein 1B #status predicted <VP2>
F:133-570/Product: coat protein 1C #status predicted <VP3>
F:571-880/Product: coat protein 1D #status predicted <VP1>
F:881-1001/Product: core protein 2A #status predicted <P2A>
F:1002-1100/Product: core protein 2B #status predicted <P2B>
F:1101-1429/Product: core protein 2C #status predicted <P2C>
F:1430-1518/Product: protein 3A #status predicted <P3A>
F:1519-1540/Product: genome-linked protein VPg #status predicted <P3>
F:1541-1723/Product: proteinase #status predicted <PC3>
F:1724-2185/Product: RNA-directed RNA polymerase #status predicted <P3>
F:1521/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 30; DB 1; Length 2185;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
Db 1425 EALFOG 1430

RESULT 9
JQ2021
genome polypeptide - coxsackievirus B5 (strain 1954/UK/85)
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; cor
polymerase (EC 2.7.7.48)
C:Species: coxsackievirus B5
C:Date: 24-Feb-1994 #sequence_revision 24-Feb-1994 #text_change 16-Jul-1999
A:Accession: JQ2021
R.Zhang, G.; Wilsden, G.; Knowles, N.J.; McCauley, J.W.

J. Gen. Virol. 74, 845-853, 1993
 A:Title: Complete nucleotide sequence of a coxsackie B5 virus and its relationship to sw
 A:Reference number: J02021; MUID:93260398; PMID:8388019
 A:Accession: J02021
 A:Molecule type: genomic RNA
 A:Residues: 1-2185 <ZHA>
 A:Cross-references: GB:X67706; NID:959045; PIDN:CAA47944.1; PID:959046
 C:Superfamily: poliovirus genome polypeptide
 C:Keywords: coat protein; core protein; genome-linked protein; nucleotidyltransferase; P
 F:1-69/Product: coat protein 1A #status predicted <C1A>
 F:70-330/Product: coat protein 1B #status predicted <C1B>
 F:331-566/Product: coat protein 1C #status predicted <C1C>
 F:566-831/Product: coat protein 1D #status predicted <C1D>
 F:852-1001/Product: core protein 2A #status predicted <C2A>
 F:1002-1100/Product: core protein 2B #status predicted <C2B>
 F:1101-1429/Product: core protein 2C #status predicted <C2C>
 F:1430-1518/Product: core protein 3A #status predicted <P3A>
 F:1519-1540/Product: genome-linked protein VPg #status predicted <VPg>
 F:1541-1723/Product: proteinase 3C #status predicted <P3C>
 F:1724-2185/Product: RNA-directed RNA polymerase #status predicted <RPS>
 F:1521/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 30; DB 1; Length 2185;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 1425 EALFOG 1430

RESULT 10

GNNYSV
 genome polypeptide - swine vesicular disease virus (strain UKG/27/72)
 N:Contains: coat protein VP1; coat protein VP2; coat protein VP3; coat protein VP4; core
 A-directed RNA polymerase (EC 2.7.7.48)
 C:Species: swine vesicular disease virus
 C:Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 16-Jul-1999
 C:Accession: S11670; A30061
 R:Seechurn, P.; Knowles, N.J.; McCauley, J.W.
 Virus Res. 16, 255-274, 1990
 A:Title: The complete nucleotide sequence of a pathogenic swine vesicular disease virus.
 A:Reference number: S11670; MUID:90364770; PMID:2168111
 A:Molecule type: genomic RNA
 A:Residues: 1-2185 <SEE>
 A:Cross-references: EMBL:X54521; NID:961167; PIDN:CAA38377.1; PID:961168
 A:Note: Submitted to the Protein Sequence Database, April 1989
 C:Superfamily: poliovirus genome polypeptide
 C:Keywords: coat protein; core protein; genome-linked protein; hydrolase; nucleotidyltra

F:1-69/Product: coat protein VP4 #status predicted <VP4>
 F:70-330/Product: coat protein VP2 #status predicted <VP2>
 F:331-566/Product: coat protein VP3 #status predicted <VP3>
 F:566-851/Product: coat protein VP1 #status predicted <VP1>
 F:852-1001/Product: core protein P2-3b #status predicted <P2A>
 F:1002-1100/Product: core protein P2-3b #status predicted <P2A>
 F:1101-1429/Product: core protein P2-X #status predicted <P2B>
 F:1430-1518/Product: core protein P3-1b #status predicted <P3A>
 F:1519-1540/Product: genome-linked protein VPg #status predicted <P3B>
 F:1541-1723/Product: proteinase 3C #status predicted <P3C>
 F:1724-2185/Product: RNA-directed RNA polymerase #status predicted <P3D>
 F:1521/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 30; DB 1; Length 2185;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 1425 EALFOG 1430

RESULT 11

GNNYSV
 genome polypeptide - swine vesicular disease virus (strain H/3/76)
 N:Contains: coat protein VP1; coat protein VP2; coat protein VP3; coat protein VP4; c
 A-directed RNA polymerase (EC 2.7.7.48)
 C:Species: swine vesicular disease virus
 C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jun-2000
 C:Accession: A31331
 R:Inoue, T.; Suzuki, K.
 J. Gen. Virol. 70, 919-934, 1989
 A:Title: The complete nucleotide sequence of swine vesicular disease virus.
 A:Reference number: A31331; MUID:89279274; PMID:2543767
 A:Accession: A31331
 A:Molecule type: genomic RNA
 A:Residues: 1-2185 <INO>
 A:Cross-references: GB:D00435; NID:91228947; PIDN:BAA00337.1; PID:9222630
 C:Superfamily: poliovirus genome polypeptide
 C:Keywords: coat protein; core protein; genome-linked protein; hydrolase; nucleotidyl
 F:1-69/Product: coat protein VP4 #status predicted <VP4>
 F:70-330/Product: coat protein VP2 #status predicted <VP2>
 F:331-566/Product: coat protein VP3 #status predicted <VP3>
 F:566-851/Product: coat protein VP1 #status predicted <VP1>
 F:852-1001/Product: core protein P2-3b #status predicted <P2A>
 F:1002-1100/Product: core protein P2-3b #status predicted <P2A>
 F:1101-1429/Product: core protein P2-X #status predicted <P2B>
 F:1430-1518/Product: core protein P3-1b #status predicted <P3A>
 F:1519-1540/Product: genome-linked protein VPg #status predicted <P3B>
 F:1541-1723/Product: proteinase 3C #status predicted <P3C>
 F:1724-2185/Product: RNA-directed RNA polymerase #status predicted <P3D>
 F:1521/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 30; DB 1; Length 2185;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 1425 EALFOG 1430

RESULT 12

S44194
 polypeptide - echovirus 12
 C:Species: echovirus 12
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999
 C:Accession: S44194
 R:Kraus, W.; Nielsen-Salzs, B.E.
 submitted to the EMBL Data Library, February 1994
 A:Reference number: S44194
 A:Accession: S44194
 A:Status: preliminary
 A:Molecule type: genomic RNA
 A:Residues: 1-2193 <KRA>
 A:Cross-references: EMBL:X77708; NID:92058304; PIDN:CAA54783.1; PID:9474421
 C:Superfamily: poliovirus genome polypeptide
 C:Keywords: genome-linked protein; phosphoprotein; polypeptide
 F:1529/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 30; DB 2; Length 2193;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 1433 EALFOG 1438

RESULT 13

S44251
 polypeptide - echovirus 12
 C:Species: echovirus 12
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 05-Jun-1998
 C:Accession: S44251
 R:Zimmermann, H.; Kraus, W.; Nielsen-Salzs, B.E.

submitted to the EMBL Data Library, February 1994

A:Reference number: S44251

A:Accession: S44251

A>Status: preliminary

A:Molecule type: genomic RNA

A:Residues: 1-2193 <ZIM>

A:Cross-references: EMBL:X79047

C:Superfamily: poliovirus genome polyprotein

C:Keywords: genome-linked protein; phosphoprotein; polyprotein

F:1529/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match

Best Local Similarity 100.0%; Score 30; DB 2; Length 2193;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6

Db 1433 EALFOG 1438

RESULT 14

S52919

polyprotein (1A, 1B, 1C, 1D, 2A, 2B, 2C, 3A, 3B, 3C, 3D) - echovirus 9

C:Species: echovirus 9

C:Date: 06-Jun-1995 #sequence-revision 21-Jul-1995 #text-change 05-Jun-1998

C:Accession: S52919

R:Zimmermann, H.; Eggers, H.J.; Nelsen-Salz, B.

submitted to the EMBL Data Library, February 1995

A:Description: Complete nucleotide sequence and construction of an infectious clone of

A:Reference number: S52919

A:Accession: S52919

A>Status: preliminary

A:Molecule type: genomic RNA

A:Residues: 1-2193 <ZIM>

A:Cross-references: EMBL:X84981

C:Superfamily: poliovirus genome polyprotein

C:Keywords: genome-linked protein; phosphoprotein; polyprotein

F:1529/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match

Best Local Similarity 100.0%; Score 30; DB 2; Length 2193;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6

Db 1433 EALFOG 1438

RESULT 15

GNNY7

genome polyprotein - human enterovirus 70 (strain J670/71)

N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core pro

NA-directed RNA polymerase (EC 2.7.7.48)

C:Species: human enterovirus 70

C:Date: 30-Jun-1993 #sequence-revision 30-Jun-1993 #text-change 05-Jun-1998

C:Accession: A36253

R:Ryan, M.D.; Jenkins, O.; Hughes, P.J.; Brown, A.; Knowles, N.J.; Booth, D.; Minor, P.D.

J. Gen. Virol. 71, 2291-2299, 1990

A:Title: The complete nucleotide sequence of enterovirus type 70: relationships with oth

A:Reference number: A36253; M01D:91037960; PMID:2172447

A:Accession: A36253

A:Molecule type: genomic RNA

A:Residues: 1-2194 <RYA>

A:Cross-references: GB:D00820

C:Superfamily: poliovirus genome polyprotein

C:Keywords: coat protein; core protein; genome-linked protein; hydrolase; nucleotidyltra

F:1-65/Product: coat protein 1A #status predicted <VP4>

F:70-319/Product: coat protein 1B #status predicted <VP2>

F:320-561/Product: coat protein 1C #status predicted <VP3>

F:562-871/Product: coat protein 1D #status predicted <VP1>

F:872-1014/Product: core protein 2A #status predicted <P2A>

F:1015-1113/Product: core protein 2B #status predicted <P2B>

F:1114-1443/Product: core protein 2C #status predicted <P2C>

F:1444-1532/Product: protein 3A #status predicted <P3A>

F:1533-1554/Product: genome-linked protein VPg #status predicted <VPg>

F:1555-1737/Product: proteinase 3C #status predicted <P3C>

F:1738-2194/Product: RNA-directed RNA polymerase #status predicted <RPP>

F:1535/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match

Best Local Similarity 100.0%; Score 30; DB 1; Length 2194;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6

Db 1439 EALFOG 1444

Search completed: May 20, 2003, 11:39:25
Job time: 1.2315 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 08:02:28 ; Search time 0.141625 Seconds

(without alignments)
1757.161 Million cell updates/sec

Title: US-10-081-408-6

Perfect score: 30

Sequence: 1 EALFOG 6

Scoring table: BLOSUM62

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	30	100.0	1081	PDR6_YEAST
2	30	100.0	2175	POLG_BOVEY
3	30	100.0	2182	POLG_CXB1J
4	30	100.0	2183	POLG_CXB4E
5	30	100.0	2183	POLG_CXB4J
6	30	100.0	2184	POLG_EC01F
7	30	100.0	2185	POLG_CXB3N
8	30	100.0	2185	POLG_CXB3W
9	30	100.0	2185	POLG_CXB3P
10	30	100.0	2185	POLG_CXB3V
11	30	100.0	2185	POLG_CXB3V
12	30	100.0	2191	POLG_EC06C
13	30	100.0	2193	POLG_CXA16
14	30	100.0	2193	POLG_EC09H
15	30	100.0	2193	POLG_EC12T
16	30	100.0	2193	POLG_HE71B
17	30	100.0	2193	POLG_HE71M
18	30	100.0	2194	POLG_EC30B
19	30	100.0	2194	POLG_EC701
20	30	100.0	2195	POLG_EC11G
21	30	100.0	2196	POLG_EC05N
22	30	100.0	2201	POLG_CXA9
23	30	100.0	2203	POLG_EC09B
24	30	100.0	2206	POLG_CXA21
25	30	100.0	2206	POLG_POL1M
26	30	100.0	2206	POLG_POL1M
27	30	100.0	2206	POLG_POL1M
28	30	100.0	2206	POLG_POL1M
29	30	100.0	2208	POLH_POL1M
30	30	100.0	2209	POLG_POL1S
31	30	100.0	2214	POLG_CXA24
32	28	93.3	2164	POLG_HRV89
33	27	90.0	272	COX3_PYL1I

ALIGNMENTS

34	27	90.0	323	1	TTD2_HUMAN	P09758	homo sapien
35	27	90.0	365	1	FXH1_HUMAN	O75593	homo sapien
36	27	90.0	395	1	SYW_AQ0AE	O67115	aquifex aeo
37	27	90.0	443	1	IF5_PHAVU	P48724	phaeosius v
38	27	90.0	452	1	HIS7_PHYPR	P28624	phyophthor
39	27	90.0	614	1	SPAT_BACSU	P33116	baecillus su
40	27	90.0	707	1	HLV2_ECOLI	P10089	escherichia
41	27	90.0	707	1	LYV5_ECOLI	P08716	escherichia
42	27	90.0	1337	1	YDM5_SCHPO	P87136	schizosacch
43	26	86.7	167	1	FTN_CAMJE	O46106	campylobact
44	26	86.7	196	1	HIS7_HA1N1	O9h13	halobacteri
45	26	86.7	201	1	CYC4_PSEAE	P00106	pseudomonas

RESULT 1

ID PDR6_YEAST STANDARD: PRT: 1081 AA.

AC P32767;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Pleiotropic drug resistance regulatory protein 6;
GN PDR6 OR YGL016W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL125-2B;
RA MEDLINE=91353083; PubMed=1882553;
RX Chen W., Balzi E., Capileau E., Choder M., Goffeau A.;
RT "The DNA sequencing of the 17 kb HindIII fragment spanning the LEU1
RT and ATR1 loci on chromosome VII from Saccharomyces cerevisiae reveals
RT the PDR6 gene, a new member of the genetic network controlling
RT pleiotropic drug resistance.";
RL Yeast 7:287-299(1991)..
RN [2]
RP SEQUENCE FROM N.A.
RA Hedling U., Hofmann B., Dellius H.;
RL Submitted (MAY 1996) to the EMBL/Genbank/DBJ databases.
CC - FUNCTION: PROBABLE POSITIVE REGULATOR OF PROTEINS INVOLVED IN
CC PERMEABILITY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: S58126; AAD13801.1;
DR EMBL: S57895; AAB19613.1;
DR EMBL: Z72538; CAA96716.1;
DR PIR: S15040; S15040.
DR SGD: S0002984; KAP122.
SQ SEQUENCE 1081 AA; 123530 MW; EEBG5ADD4E16D9D3 CRC64;

Query Match Best Local Similarity 100.0%; Score 30; DB 1; Length 1081;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
DB 631 EALFOG 636

RESULT 2
POLG_BOVEY

1D POLG_BOVEV STANDARD; PRT; 2175 AA.
 AC P12915;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Genome polypeptide [Contains: Coat proteins VP1 TO VP4; Core proteins
 P2A TO P2C, P3A; Genome-linked protein VPg; Plicorinain 3C
 DE (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase P3D
 DE (EC 2.7.7.48)]
 OS Bovine enterovirus (strain VG-5-27) (BEV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.
 OX NCBI_TaxID=12065;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88117392; PubMed=2828511;
 RA Earle J.A.P., Skuce R.A., Fleming C.S., Hoey E.M., Martin S.J.;
 RT "The complete nucleotide sequence of a bovine enterovirus.";
 RL J. Gen. Virol. 69:253-263(1988).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF 1-840.
 RX MEDLINE=95292108; PubMed=7773791;
 RA Smyth M., Tate J., Hoey E.M., Lyons C., Martin S.J., Stuart D.;
 RT "Implications for viral uncoating from the structure of bovine
 RT enterovirus.";
 RL Nat. Struct. Biol. 2:224-231(1995).
 CC -1- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
 CC Q/G SITES IN THE POLYPEPTIDE. IT MAY BE A CYSTEINE PROTEASE.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Gln-1-Gly bond in the
 CC poliovirus polypeptide. In other picornavirus reactions Glu may be
 CC substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC [RNA](N).
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: D00214; BAA24003.1; ALT_SEQ.
 DR PIR: A29824; GNNYBE.
 DR PDB: 1BEV; 16-SEP-98.
 DR MEROPS: C03 UPA; -;
 DR MEROPS: C03 UPA; -;
 DR InterPro: IPR000199; Cys-protease-3C.
 DR InterPro: IPR000318; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR000605; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR InterPro: IPR001676; RNv.
 DR Pfam: PF00073; RNv; 3.
 DR Pfam: PF00548; Cys-protease-3C; 1.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
 DR Pfam: PF00910; RNA_helicase; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR ProDom: PD001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 KM Polypeptide: Coat protein; Core protein; Transferase;
 KM RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate;
 KM 3D-structure.
 FT CHAIN 2 69 COAT PROTEIN VP4 (P1A).

FT CHAIN 70 317 COAT PROTEIN VP2 (P1B).
 FT CHAIN 318 559 COAT PROTEIN VP3 (P1C).
 FT CHAIN 560 840 COAT PROTEIN VP4 (P1D).
 FT CHAIN 841 990 CORE PROTEIN P2A.
 FT CHAIN 991 1088 CORE PROTEIN P2B.
 FT CHAIN 1090 1419 CORE PROTEIN P2C.
 FT CHAIN 1420 1508 CORE PROTEIN P3A.
 FT CHAIN 1509 1531 GENOME-LINKED PROTEIN VPg (P3B).
 FT CHAIN 1532 1714 PICORINAIN 3C.
 FT CHAIN 1715 2175 RNA-DEPENDENT RNA POLYMERASE P3D.
 FT LIPID 2 2175 MYRISTATE.
 FT ACT_SITE 1678 1678 PROTEASE (POTENTIAL).
 FT ACT_SITE 1692 1692 PROTEASE (POTENTIAL).
 SQ SEQUENCE 2175 AA; 242502 MW; 44FCAD8704E48FD CRC64;
 OY 1 EALFOG 6
 Db 1415 EALFOG 1420
 OY 1 EALFOG 6
 Db 1415 EALFOG 1420
 RESULT 3
 ID POLG_CXB1J STANDARD; PRT; 2182 AA.
 AC P08291;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Genome polypeptide [Contains: Coat protein VP4 (P1A); Coat protein VP2
 DE (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Plicorinain 2A
 DE (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein
 DE P3A; Genome-linked protein VPg (P3B); Plicorinain 3C (EC 3.4.22.28)
 DE (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
 OS Coxsackievirus B1 (strain Japan).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.
 OX NCBI_TaxID=103902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87122157; PubMed=3027969;
 RA Itzuka N., Kuge S., Nomoto A.;
 RT "Complete nucleotide sequence of the genome of coxsackievirus B1.";
 RL Virology 156:64-73(1987).
 CC -1- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE
 CC AT CERTAIN Q/G SITES IN THE POLYPEPTIDE. THEY ARE CYSTEINE
 CC PROTEASES.
 CC -1- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR
 CC MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Gln-1-Gly bond in the
 CC poliovirus polypeptide. In other picornavirus reactions Glu may be
 CC substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Tyr-1-Gly bond in the
 CC picornavirus polypeptide. In other picornavirus reactions Glu may
 CC be substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC [RNA](N).
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC: VP1/VP2A IS
 CC CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
 CC -1- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

```

CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to licenselsib@sib.ch).
CC  -----
DR  EMBL; M16560; AAC00531.1; -.
DR  PIR; A26353; GNMYB1.
DR  HSSP; P21404; IDAM.
DR  MEROPS; C03.011; -.
DR  MEROPS; C03.022; -.
DR  InterPro; IPR000199; Cys-protease-3C.
DR  InterPro; IPR003138; Pico_P1A.
DR  InterPro; IPR000081; Pico_P2A.
DR  InterPro; IPR002527; Pico_P2B.
DR  InterPro; IPR000605; RNA_helicase.
DR  InterPro; IPR001205; RNA_pol_P3D.
DR  InterPro; IPR001676; Rny.
DR  Pfam; PF000073; Rny; 3.
DR  Pfam; PF00548; Cys-protease-3C; 1.
DR  Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR  Pfam; PF00910; RNA_helicase; 1.
DR  Pfam; PF00947; Pico_P2A; 1.
DR  Pfam; PF01552; Pico_P2A; 1.
DR  Pfam; PF02226; Pico_P1A; 1.
DR  ProDom; PD001125; Cys-protease-3C; 1.
DR  ProDom; PD001274; Pico_P2B; 1.
DR  ProDom; PD001306; Pico_P2A; 1.
KW  Polyprotein; Coat protein; Core protein; Transferase; Myristate;
KW  RNA-directed RNA polymerase; Hydroxylase; Thiol protease.
FT  CHAIN 2 69
FT  CHAIN 70 332
FT  CHAIN 333 570
FT  CHAIN 571 848
FT  CHAIN 849 998
FT  CHAIN 999 1097
FT  CHAIN 1098 1426
FT  CHAIN 1427 1515
FT  CHAIN 1516 1537
FT  CHAIN 1538 1720
FT  CHAIN 1721 2182
FT  LIPID 2
FT  ACT_SITE 1684 1684
FT  ACT_SITE 1698 1698
SQ  SEQUENCE 2182 AA; 243945 MW; 387B9391275859B1 CRC64;

Query Match 100.0%; Score 30; DB 1; Length 2182;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
DB 1422 EALFOG 1427

RESULT 4
POLG_CXB4E
ID POLG_CXB4E STANDARD; PRT; 2183 AA.
AC 086887;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contents: Coat protein VP4 (P1A); Coat protein VP2
DE (P1B); Coat protein VP3 (P1C); Coat protein VP4 (P1A); Coat protein VP2
DE (EC 3.4.22.39) (P2A); Core protein P2B; Core protein P2C; Core protein
DE P3A; Genome-linked protein VP6 (P3B); Picornain 3C (EC 3.4.22.28)
DE (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
OS Coxsackievirus B4 (strain E2).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Enterovirus.
OX NCBI_TaxID=103905;
RN 11
RP SEQUENCE FROM N.A.
RA MEDLINE=95205102; PubMed=7897366;
RA Kang Y., Chatterjee N.K., Nodwell M.J., Yoon J.W.;
RT Complete nucleotide sequence of a strain of coxsackie B4 virus of

```

Query	Match	Score	DB	Length
RT	human origin that induces diabetes in mice and its comparison with	100.0%	Score30	DB 1
RT	nonabetogenic coxsackie B4 JBV strain."	100.0%	Score30	DB 1
RL	J. Med. Virol. 44:353-361(1994).	100.0%	Score30	DB 1
CC	-1- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE	100.0%	Score30	DB 1
CC	AT CERTAIN Q/G SITES IN THE POLYPROTEIN. THEY ARE CYSTEINE	100.0%	Score30	DB 1
CC	PROTEASES.	100.0%	Score30	DB 1
CC	-1- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR	100.0%	Score30	DB 1
CC	MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.	100.0%	Score30	DB 1
CC	-1- CATALYTIC ACTIVITY: Selective cleavage of gln-1-gly bond in the	100.0%	Score30	DB 1
CC	polyovirus polyprotein. In other picornavirus reactions Glu may be	100.0%	Score30	DB 1
CC	substituted for Glu, and Ser or Thr for Gly.	100.0%	Score30	DB 1
CC	-1- CATALYTIC ACTIVITY: Selective cleavage of Tyr-1-gly bond in the	100.0%	Score30	DB 1
CC	picornavirus polyprotein. In other picornavirus reactions Glu may	100.0%	Score30	DB 1
CC	be substituted for Glu, and Ser or Thr for Gly.	100.0%	Score30	DB 1
CC	-1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +	100.0%	Score30	DB 1
CC	(RNA)(N).	100.0%	Score30	DB 1
CC	-1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,	100.0%	Score30	DB 1
CC	EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,	100.0%	Score30	DB 1
CC	VP3, AND VP4.	100.0%	Score30	DB 1
CC	-1- PRIM: SPECIFIC ENZYMAITC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.	100.0%	Score30	DB 1
CC	CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC; VP1/PA IS	100.0%	Score30	DB 1
CC	CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.	100.0%	Score30	DB 1
CC	-1- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.	100.0%	Score30	DB 1
CC	-1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.	100.0%	Score30	DB 1
CC	-----	100.0%	Score30	DB 1
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	100.0%	Score30	DB 1
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	100.0%	Score30	DB 1
CC	the European Bioinformatics Institute. There are no restrictions on its	100.0%	Score30	DB 1
CC	use by non-profit institutions as long as its content is in no way	100.0%	Score30	DB 1
CC	modified and this statement is not removed. Usage by and for commercial	100.0%	Score30	DB 1
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	100.0%	Score30	DB 1
CC	or send an email to license@isb-sib.ch).	100.0%	Score30	DB 1
CC	-----	100.0%	Score30	DB 1
DR	EMBL: S76772; AAB33885.1; -	100.0%	Score30	DB 1
DR	HSSP: P21404; ID4M.	100.0%	Score30	DB 1
DR	Interpro: IPR004004; Calicl-pol_hel.	100.0%	Score30	DB 1
DR	Interpro: IPR000199; Cys-protease-3C.	100.0%	Score30	DB 1
DR	Interpro: IPR003138; Pico_P1A.	100.0%	Score30	DB 1
DR	Interpro: IPR000081; Pico_P2A.	100.0%	Score30	DB 1
DR	Interpro: IPR000527; Pico_P2B.	100.0%	Score30	DB 1
DR	Interpro: IPR000605; RNA_helicase.	100.0%	Score30	DB 1
DR	Interpro: IPR001205; RNA_pol_P3D.	100.0%	Score30	DB 1
DR	Interpro: IPR001676; Rnv.	100.0%	Score30	DB 1
DR	Pfam: PF00073; Ihv; 3.	100.0%	Score30	DB 1
DR	Pfam: PF00548; Cys-protease-3C; 1.	100.0%	Score30	DB 1
DR	Pfam: PF00680; RNA_dep_RNA_pol; 1.	100.0%	Score30	DB 1
DR	Pfam: PF00910; RNA_helicase; 1.	100.0%	Score30	DB 1
DR	Pfam: PF00947; Pico_P2A; 1.	100.0%	Score30	DB 1
DR	Pfam: PF01552; Pico_P2B; 1.	100.0%	Score30	DB 1
DR	Pfam: PF02226; Pico_P1A; 1.	100.0%	Score30	DB 1
DR	PRINTS; PR00918; CALICIVIRUS.	100.0%	Score30	DB 1
DR	ProDom: PD001125; Cys-protease-3C; 1.	100.0%	Score30	DB 1
DR	ProDom: PD001274; Pico_P2B; 1.	100.0%	Score30	DB 1
DR	ProDom: PD001306; Pico_P2A; 1.	100.0%	Score30	DB 1
DR	ProDom: PD001306; Pico_P2B; 1.	100.0%	Score30	DB 1
KM	Polyprotein; Coat protein; Core protein; Transferase; Myristate;	100.0%	Score30	DB 1
KM	RNA-directed RNA polymerase; Hydrolyase; Thiol protease.	100.0%	Score30	DB 1
FT	CHAIN 2 69 COAT PROTEIN VP4.	100.0%	Score30	DB 1
FT	CHAIN 70 330 COAT PROTEIN VP2.	100.0%	Score30	DB 1
FT	CHAIN 331 568 COAT PROTEIN VP3.	100.0%	Score30	DB 1
FT	CHAIN 569 849 COAT PROTEIN VP1.	100.0%	Score30	DB 1
FT	CHAIN 850 999 PICORNAIN 2A.	100.0%	Score30	DB 1
FT	CHAIN 1000 1098 CORE PROTEIN P2B.	100.0%	Score30	DB 1
FT	CHAIN 1099 1427 CORE PROTEIN P2C.	100.0%	Score30	DB 1
FT	CHAIN 1428 1516 CORE PROTEIN P3A.	100.0%	Score30	DB 1
FT	CHAIN 1517 1538 GENOME-LINKED PROTEIN VPG.	100.0%	Score30	DB 1
FT	CHAIN 1539 1721 PICORNAIN 3C.	100.0%	Score30	DB 1
FT	CHAIN 1722 2163 RNA-D			

Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
DB 1423 EALFOG 1428

RESULT 5
POLG_CXB4J STANDARD; PRT; 2183 AA.
ID POLG_CXB4J
AC P08292;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Coat protein VP1 (P1A); Coat protein VP2 (P1B); Coat protein VP3 (P1C); Coat protein VP4 (P1D); Picornain 2A (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein P3A; Genome-linked protein VPg (P3B); Picornain 3C (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
DE Coxsackievirus B4 (strain JVB / Benschooten / New York/51).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae; Enterovirus.
OC NCBI_TaxID=103906;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87253111; PubMed=3037008;
RA Jenkins O., Booth J.D., Minor P.D., Almond J.W.;
RT "The complete nucleotide sequence of coxsackievirus B4 and its comparison to other members of the Picornaviridae.";
RL J. Gen. Virol. 68:1835-1848(1987).
RN [2]
RP REVISIONS.
RA Jenkins O.;
RL Submitted (AUG-1989) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE AT CERTAIN Q/G SITES IN THE POLYPEPTIDE. THEY ARE CYSTEINE PROTEASES.
CC -1- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of Gln-1-Gly bond in the poliovirus polyprotein. In other picornavirus reactions Gln may be substituted for Gln, and Ser or Thr for Gly.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of Tyr-1-Gly bond in the picornavirus polyprotein. In other picornavirus reactions Gln may be substituted for Gln, and Ser or Thr for Gly.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate + [RNA](N).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS, EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.
CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS. CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC; VP1/P2A IS CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
CC -1- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X05690; CAA29172.1; -;
DR PIR: A27170; GNNYB4.
DR HSSP: P21404; 1D4M.
DR MEROPS: C03.011; -;
DR MEROPS: C03.022; -;
DR InterPro: IPR000199; Cys-protease-3C.
DR InterPro: IPR003138; Pico_P1A.
DR InterPro: IPR000081; Pico_P2A.

DR InterPro: IPR002527; Pico_P2B.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR InterPro: IPR001676; Rnv.
DR Pfam: PF00073; rhv; 3.
DR Pfam: PF00548; Cys-protease-3C; 1.
DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR Pfam: PF00910; RNA_helicase; 1.
DR Pfam: PF00947; Pico_P2A; 1.
DR Pfam: PF01552; Pico_P2B; 1.
DR Pfam: PF02226; Pico_P1A; 1.
DR ProDom: PD001125; Cys-protease-3C; 1.
DR ProDom: PD001274; Pico_P2B; 1.
DR ProDom: PD001306; Pico_P2A; 1.
KM Polyprotein: Coat protein; Core protein; Transferase; Myristate;
FT RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 2
FT CHAIN 69
FT CHAIN 70 330
FT CHAIN 331 568
FT CHAIN 569 849
FT CHAIN 850 999
FT CHAIN 1000 1098
FT CHAIN 1099 1427
FT CHAIN 1428 1516
FT CHAIN 1517 1538
FT CHAIN 1539 1721
FT CHAIN 1722 2183
FT CHAIN 2
FT LIPID 2
FT ACT_SITE 1685 1685
FT ACT_SITE 1699 1699
FT ACT_SITE 1699 1699
SQ SEQUENCE 2183 AA; 244012 MW; 244012 MW; DE069DE3AE91AF0 CRC64;

Query Match Best Local Similarity 100.0%; Score 30; DB 1; Length 2183;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
DB 1423 EALFOG 1428

RESULT 6
POLG_EC01F STANDARD; PRT; 2184 AA.
ID POLG_EC01F
AC 091734; 066795; 09YID6;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Genome polyprotein [Contains: Coat protein VP4 (P1A); Coat protein VP2 (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Picornain 2A (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein P3A; Genome-linked protein VPg (P3B); Picornain 3C (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
DE Echovirus 1 (strain Farok / ATCC VR-1038).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae; Enterovirus.
OC NCBI_TaxID=103908;
RN [1]
RP SEQUENCE FROM N.A.
RA Bergelson J.M.;
RT "Receptor interactions, infectious cDNA, and nucleotide sequences of echovirus 1/8.";
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 69-330 FROM N.A.
RX MEDLINE=96203959; PubMed=8627260;
RA Hutunen P., Santti J., Pulli T., Hyypia T.;
RT "The major echovirus group is genetically coherent and related to coxsackie B viruses.";
RL J. Gen. Virol. 77:715-725(1996).
RN [3]
RP SEQUENCE OF 570-853 FROM N.A.
RX MEDLINE=99138973; PubMed=9971773;

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 CC -1- PPM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC. VP1/VP2 IS
 CC CATALYZED BY P2A. ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
 CC -1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M33854; AAA42931.1; -;
 DR EMBL: K02709; AAA42932.1; -;
 DR EMBL: M16572; AAA74400.1; -;
 DR PIR: A26354; GNNYB3.
 DR PIR: A34664; GNNYBT.
 DR PDB: 1COV; 06-MAR-96.
 DR MEROPS: C03.011; -;
 DR MEROPS: C03.022; -;
 DR InterPro: IPR000199; Cys-protease-3C.
 DR InterPro: IPR000318; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR000605; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR InterPro: IPR001676; RNv.
 DR Pfam: PF00073; tnv; 3.
 DR Pfam: PF00548; Cys-protease-3C; 1.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
 DR Pfam: PF00910; RNA_helicase; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR ProDom: PD001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 KW Polyprotein; Coat protein; Core protein; Transferase; Myristate;
 KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; 3D-structure.
 KM RNA-directed RNA polymerase: Hydrolase; Thiol protease; 3D-structure.
 FT CHAIN 2
 FT CHAIN 69
 FT CHAIN 70 332
 FT CHAIN 333 570
 FT CHAIN 571 851
 FT CHAIN 852 1001
 FT CHAIN 1002 1100
 FT CHAIN 1101 1429
 FT CHAIN 1430 1518
 FT CHAIN 1519 1540
 FT CHAIN 1541 1723
 FT CHAIN 1724 2185
 FT CHAIN 2185
 FT LIPID 2
 FT ACT_SITE 1687
 FT ACT_SITE 1701
 FT CONFLICT 16 16
 FT CONFLICT 17 17
 FT CONFLICT 469 469
 FT CONFLICT 487 487
 FT CONFLICT 510 510
 FT CONFLICT 516 516
 FT CONFLICT 566 566
 FT CONFLICT 593 593
 FT CONFLICT 650 650
 FT CONFLICT 854 865
 FT CONFLICT 873 873
 FT CONFLICT 1097 1097
 FT CONFLICT 1280 1280

FT CONFLICT 1437 1437 I -> F (IN REF. 2).
 FT CONFLICT 1503 1503 V -> M (IN REF. 2).
 FT CONFLICT 1616 1616 K -> E (IN REF. 2).
 FT CONFLICT 1624 1624 R -> G (IN REF. 2).
 FT CONFLICT 1627 1627 R -> G (IN REF. 2).
 FT CONFLICT 1630 1630 L -> V (IN REF. 2).
 FT CONFLICT 1718 1718 Y -> N (IN REF. 2).
 FT CONFLICT 1734 1734 D -> V (IN REF. 2).
 FT CONFLICT 1758 1758 E -> V (IN REF. 2 AND 3).
 FT CONFLICT 1824 1824 V -> R (IN REF. 2 AND 3).
 FT CONFLICT 1867 1867 C -> R (IN REF. 2 AND 3).
 FT CONFLICT 1880 1880 Y -> H (IN REF. 2 AND 3).
 FT CONFLICT 2001 2001 D -> N (IN REF. 2 AND 3).
 FT CONFLICT 2095 2095 A -> V (IN REF. 2 AND 3).
 FT CONFLICT 2115 2115 V -> A (IN REF. 2 AND 3).
 FT CONFLICT 2175 2175 S -> T (IN REF. 2 AND 3).
 FT CONFLICT 2178 2178 R -> G (IN REF. 2 AND 3).
 SQ SEQUENCE 2185 AA; 243451 MW; 1B5CE3DA7338FF CRC64;
 Query Match 100.0%; Score 30; DB 1; Length 2185;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 BALFOG 6
 DB 1425 BALFOG 1430
 ID POLG_CXB3M STANDARD: PRT: 2185 AA.
 AC Q66282; 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Coat protein VP4 (P1A); Coat protein VP2
 DE (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Picornain 2A
 DE (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein
 DE P3A; Genome-linked protein VPg (P3B); Picornain 3C (EC 3.4.22.28)
 DE (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
 OS Coxsackievirus B3 (strain Woodruff).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.
 OX NCBI_TaxID=103904;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Knowlton K.U., Jeon E.S., Berkley R.W., Messely R., Huber S.;
 RT "A mutation in the puff region of VP2 attenuates the myocarditic
 RT phenotype of an infectious cDNA of the Woodruff virus";
 RL Submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE
 CC AT CERTAIN Q/G SITES IN THE POLYPROTEIN. THEY ARE CYSTEINE
 CC PROTEASES.
 CC -1- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR
 CC MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Gln-1-gly bond in the
 CC poliovirus polyprotein. In other picornavirus reactions Glu may be
 CC substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Tyr-1-gly bond in the
 CC picornavirus polyprotein. In other picornavirus reactions Glu may
 CC be substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 CC -1- PPM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC. VP1/VP2 IS
 CC CATALYZED BY P2A. ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
 CC -1- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

Query Match	100.0%	Score 30	DB 1	Length 2185
Best Local Similarity	100.0%	Pred. No. 52		
Matches	6	Conservative 0	Mismatches 0	Indels 0
				Gaps 0

RESULT 9	
POLG_CXB5P	
ID	STANDARD;
00000000	PRT; 2185 AA

DT 01-JUL-1993 (Rel. 26, last sequence update)
 DT 15-JUN-2002 (Rel. 41, last annotation update)
 DE Genome polyprotein [contains: Coat protein VP4 (P1A); Coat protein VP7
 DE (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Pliconin 2A
 DE (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein
 DE P3A; Genome-linked protein VP6 (P3B); Pliconin 3C (EC 3.4.22.28)
 DE (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)]
 OC Coxsackievirus B5 (strain Peterborough / 1954/PK/85).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.
 NCBI_TaxID=103907;

[1]
 SEQUENCE FROM N.A.. PubMed-8388019;
 MEDLINE-93260398; Zhang G., Wilsden G., Knowles N.J., McCauley J.W.;
 "Complete nucleotide sequence of a coxsackie B5 virus and its
 relationship to swine vesicular disease virus.";
 J. Gen. Virol. 74:845-853(1993).
 -1- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE
 AT CERTAIN Q/G SITES IN THE POLYPROTEIN. THEY ARE CYSTEINE
 PROTEASES.
 -1- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR
 MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.
 -1- CATALYTIC ACTIVITY: Selective cleavage of Glu-1-gly bond in the
 poliovirus polyprotein. In other picornavirus reactions Glu may be
 substituted for Gln, and Ser or Thr for Gly.
 -1- CATALYTIC ACTIVITY: selective cleavage of Tyr-1-gly bond in the
 picornavirus polyprotein. In other picornavirus reactions Glu may
 be substituted for Gln, and Ser or Thr for Gly.
 -1- CATALYTIC ACTIVITY: N nucleoside triphosphate +
 {RNA}(N).
 -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
 EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 VP3, AND VP4.
 -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC; VP1/P2A IS
 CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
 -1- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 -1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb.ch/announce/>
 or send an email to license@isb.sib.ch).

 EMBL: X67706; CAA47944.1; -
 DR PIR: J02021; J02021.
 DR HSP: P21404; ID4M.
 DR MEROPS: C03.011; -
 DR MEROPS: C03.022; -
 DR InterPro: IPR000109; Cys-protease-3C.
 DR InterPro: IPR003138; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR000605; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR InterPro: IPR001676; Rhv.
 DR Pfam: PF000073; Rhv_3.
 DR Pfam: PF00548; Cys-protease-3C; 1.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
 DR Pfam: PF00910; RNA_helicase; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01553; Pico_P2B; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR ProDom: PD0001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 KW Polypeptide; Coat protein; Core protein; Core protein; Transferase; Myristate;
 KM RNA-directed RNA polymerase; Hydrolyase; Ribol protease.
 DR CHAIN 2 69
 FT CHAIN 70 330
 FT CHAIN 331 568
 FT CHAIN 569 851
 FT CHAIN 852 1001
 FT CHAIN 1002 1100
 FT CHAIN 1101 1429
 FT CHAIN 1430 1518
 FT CHAIN 1519 1540
 FT CHAIN 1541 1723
 FT CHAIN 1724 2185
 FT CHAIN 2 2
 LPID 2
 MYRISTATE (BY SIMILARITY).
 P1CORNAIN 3C.
 RNA-DIRECTED RNA POLYMERASE.
 P1CORNAIN 3C.
 MYRISTATE (BY SIMILARITY).

FT ACT_SITE 1687 1687 PROTEASE 3C (POTENTIAL).
 FT ACT_SITE 1701 1701 PROTEASE 3C (POTENTIAL).
 SO SEQUENCE 2185 AA; 243298 MW; 3F9EE29F90D59C6F CRC64;

Query Match 100.0%; Score 30; DB 1; Length 2185;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 1425 EALFOG 1430

RESULT 10

POLG_SVDVH STANDARD; PRT; 2185 AA.

AC P16604;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Genome polypeptide (Contains: Coat proteins VP1 to VP4; Core proteins
 P2A to P2C; P3A, Genome-linked protein VP6; Picornain 3C
 (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase P3D
 (EC 2.7.7.48)).

OC Swine vesicular disease virus (strain H/3 '76).
 OC Viruses: ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.

OX NCBI_Taxid=12076;

RN [1]

RP MEDLINE=89279274; PubMed=2543767;

RA Inoue T., Suzuki T., Sekiguchi K.;

RT "The complete nucleotide sequence of swine vesicular disease virus.";

RL J. Gen. Virol. 70:919-934(1989).

-1- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR
 MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.

-1- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
 O/G SITES IN THE POLYPEPTIDE. IT MAY BE A CYSTEINE PROTEASE.

-1- CATALYTIC ACTIVITY: Selective cleavage of Gln-I-Gly bond in the
 poliovirus polypeptide. In other picornavirus reactions Glu may be
 substituted for Gln, and Ser or Thr for Gly.

-1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 (RNA)(N).

-1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
 EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 VP3, AND VP4.

-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

-1- PRT: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

-1- PRT: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

-1- PRT: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

-1- PRT: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

-1- PRT: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

-1- PRT: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

-1- PRT: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

-1- PRT: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

-1- PRT: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

-1- PRT: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR ProDom: PD001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 KM Polypeptide: Coat protein; Core protein; Transferase;
 KM RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate.

FT CHAIN 2
 FT CHAIN 70 330
 FT CHAIN 331 568
 FT CHAIN 569 851
 FT CHAIN 852 1001
 FT CHAIN 1002 1100
 FT CHAIN 1101 1429
 FT CHAIN 1430 1518
 FT CHAIN 1519 1540
 FT CHAIN 1541 1723
 FT CHAIN 1724 2185

FT LIPID 2
 FT ACT_SITE 1687 1687
 FT ACT_SITE 1701 1701

FT SO SEQUENCE 2185 AA; 243164 MW; 5B2285D9482B3B6 CRC64;

Query Match 100.0%; Score 30; DB 1; Length 2185;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 1425 EALFOG 1430

RESULT 11

POLG_SVDVH STANDARD; PRT; 2185 AA.

AC P13300; 084794; 084795; 084796; 084797; 084798; 084799; 084800;

AC 084801; 084802; 084803; 084804;

DT 01-JAN-1990 (Rel. 13, Created)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Genome polypeptide (Contains: Coat proteins VP1 to VP4; Core proteins
 P2A to P2C; P3A, Genome-linked protein VP6; Picornain 3C
 (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase P3D
 (EC 2.7.7.48)).

OC Swine vesicular disease virus (strain UKG/27/72).

OC Viruses: ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.

OX NCBI_Taxid=12077;

RN [1]

RP MEDLINE=90364770; PubMed=2168111;

RA Seechurn P., Knowles N.J., McCauley J.W.;

RT "The complete nucleotide sequence of a pathogenic swine vesicular
 disease virus.";

RL Virus Res. 16:255-274(1990).

-1- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR
 MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.

-1- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
 O/G SITES IN THE POLYPEPTIDE. IT MAY BE A CYSTEINE PROTEASE.

-1- CATALYTIC ACTIVITY: Selective cleavage of Gln-I-Gly bond in the
 poliovirus polypeptide. In other picornavirus reactions Glu may be
 substituted for Gln, and Ser or Thr for Gly.

-1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 (RNA)(N).

-1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
 EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 VP3, AND VP4.

-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

-1- PRT: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

-1- PRT: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

-1- PRT: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

-1- PRT: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL, X54521, CAA38377.1; -
 DR PIR, A30061, GNMISV.
 DR PIR, S11670, S11670.
 DR MEROPS, C03.011; -
 DR MEROPS, C03.022; -
 DR InterPro: IPR000199; Cys-protease-3C.
 DR InterPro: IPR003138; Pico_P2A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR000605; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR InterPro: IPR001676; Rnv.
 DR Pfam: PF00073; Rnv; 3.
 DR Pfam: PF00548; Cys-protease-3C; 1.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
 DR Pfam: PF00910; RNA_helicase; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF02326; Pico_P1A; 1.
 DR ProDom: PD001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 DR PolyProtein; Coat protein; Core protein; Transferrase;
 KM RNA-directed RNA polymerase; Hydrolyase; Thiol protease; Myristate.
 FT CHAIN 2 69
 FT CHAIN 70 330
 FT CHAIN 331 568
 FT CHAIN 359 851
 FT CHAIN 852 1001
 FT CHAIN 1002 1100
 FT CHAIN 1101 1429
 FT CHAIN 1430 1518
 FT CHAIN 1519 1540
 FT CHAIN 1541 1723
 FT CHAIN 1724 2185
 FT LIPID 2
 FT ACT SITE 1687 1687
 FT ACT SITE 1701 1701
 FT ACT SITE 243363 MM; C9B103052934E1B8 CRC64;
 SQ SEQUENCE 2185 AA; 243363 MM; C9B103052934E1B8 CRC64;

Query Match 100.0%; Score 30; DB 1; Length 2185;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 DB 1425 EALFOG 1430

RESULT 12
 ID POLG_EC06C STANDARD; PRT; 2191 AA.
 AC 066474;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Genome polyprotein (contains: Coat protein VP4 (P1A); Coat protein VP2
 DE (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Picornain 2A
 DE (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein
 DE P3A; Genome-linked protein VP6 (P3B); Picornain 3C (EC 3.4.22.28)
 DE (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)).
 OS Echovirus 6 (strain Charles).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.
 NC NCBL_TaxID-103913;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Rightland V.F., Gratsch T.E., Blackburn R.V.;
 RT "The genome of echovirus 6."
 RL Submitted (OCT-1994) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1526-2191 FROM N.A.
 RX MEDLINE-94240818, PubMed-8184543;
 RA Gratsch T.E., Rightland V.F.;
 RT "Constitution of a recombinant cDNA of echovirus 6 that established a
 RT persistent in vitro infection."
 RL Virology 201:341-348(1994).
 CC -1- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE
 CC AT CERTAIN Q/G SITES IN THE POLYPROTEIN. THEY ARE CYSTEINE
 CC PROTEASES.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Gln-1-Gly bond in the
 CC poliovirus polyprotein. In other picornavirus reactions Glu may be
 CC substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Tyr-1-Gly bond in the
 CC picornavirus polyprotein. In other picornavirus reactions Glu may
 CC be substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC; VP1/P2A IS
 CC CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
 CC -1- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL, U16283; AAA65044.1; -
 DR HSSP: P21404; 1D4M.
 DR MEROPS, C03.011; -
 DR InterPro: IPR004004; Calici_pol_hel.
 DR InterPro: IPR000189; Cys-protease-3C.
 DR InterPro: IPR003138; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR000605; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR InterPro: IPR001676; Rnv.
 DR Pfam: PF00073; Rnv; 3.
 DR Pfam: PF00548; Cys-protease-3C; 1.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
 DR Pfam: PF00910; RNA_helicase; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR PRINTS: PR00918; CALICIVIRUSNS.
 DR ProDom: PD001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 KM Polyprotein; Coat protein; Core protein; Transferrase; Myristate;
 KM RNA-directed RNA polymerase; Hydrolyase; Thiol protease.
 FT CHAIN 2 69
 FT CHAIN 70 330
 FT CHAIN 331 568
 FT CHAIN 359 857
 FT CHAIN 858 1007
 FT CHAIN 1008 1106
 FT CHAIN 1107 1435
 FT CHAIN 1436 1524
 FT CHAIN 1525 1546
 FT CHAIN 1547 1729
 FT CHAIN 1730 2191

RNA-DIRECTED RNA POLYMERASE.

FT LIPID 2 2 MYRISTATE (BY SIMILARITY).
 FT ACT_SITE 1693 1693 PROTEASE 3C (POTENTIAL).
 FT ACT_SITE 1707 1707 PROTEASE 3C (POTENTIAL).
 SQ SEQUENCE 2191 AA; 245275 MW; 47E0448B05716DB CRC64;

Query Match
 Best Local Similarity 100.0%; Score 30; DB 1; Length 2191;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 EALFOG 6
 Db 1431 EALFOG 1436

RESULT 13
 POLG_CXA16 STANDARD; PRT; 2193 AA.

AC O65900;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Genome polypeptide [Contains: Coat protein VP4 (P1A); Coat protein VP2 (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Core protein P2A; Core protein P2B; Core protein P2C; Core protein P3A; Genome-linked protein VP6 (P3B); Picornain 3C (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
 DE Picornavirus VP6 (strain G-10).
 OS Picornaviridae; ssRNA positive-strand viruses, no DNA stage; Picornaviridae; Enterovirus.

NCBI_TaxID=69159;
 RN (1)
 RP SEQUENCE FROM N.A.

RA MEDLINE=94303216; Pubmed=8030260;
 RA Poyry T., Hyypia T., Horsnell C., Kinnunen L., Hovi T., Stanway G.;

RT "Molecular analysis of coxsackievirus A16 reveals a new genetic group of enteroviruses."

RL Virology 202;982-987(1994).

-1- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS

-1- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN O/G SITES IN THE POLYPEPTIDE. IT MAY BE A CYSTEINE PROTEASE.

-1- CATALYTIC ACTIVITY: Selective cleavage of Gln-1-gly bond in the poliovirus polypeptide. In other picornavirus reactions Glu may be substituted for Gln, and Ser or Thr for Gly.

-1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate + (RNA)(N).

-1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS, EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.

-1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC EMBL; U05876; AA50478.1; -
 DR HSSP; P03300; IPOV.
 DR MEROPS; C03.022; -
 DR MEROPS; C03.024; -
 DR InterPro; IPR000199; Cys-protease-3C.
 DR InterPro; IPR00138; Pico_P1A.
 DR InterPro; IPR000081; Pico_P2B.
 DR InterPro; IPR002527; Pico_P2B.
 DR InterPro; IPR000605; RNA_helicase.
 DR InterPro; IPR001205; RNA_pol_P3D.
 DR InterPro; IPR001676; RNV.
 DR Pfam; PF00073; Hvi_3.
 DR Pfam; PF00548; Cys-protease-3c; 1.

DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
 DR Pfam; PF00910; RNA_helicase; 1.
 DR Pfam; PF00947; Pico_P2A; 1.
 DR Pfam; PF01552; Pico_P2B; 1.
 DR Pfam; PF02226; Pico_P1A; 1.
 DR ProDom; PD001125; Cys-protease-3C; 1.
 DR ProDom; PD001274; Pico_P2B; 1.
 DR ProDom; PD001306; Pico_P2A; 1.

KW Polyprotein; Coat protein; Core protein; Transferase; Hydrolyase; Thiol protease; Myristate.
 RN RNA-directed RNA polymerase; RNA polymerase; Myristate.
 FT CHAIN 2
 FT CHAIN 70 323
 FT CHAIN 324 565
 FT CHAIN 566 862
 FT CHAIN 863 1012
 FT CHAIN 1013 1111
 FT CHAIN 1112 1440
 FT CHAIN 1441 1526
 FT CHAIN 1527 1548
 FT CHAIN 1549 1731
 FT CHAIN 1732 2193

FT LIPID 2
 FT ACT_SITE 1695 1695
 FT ACT_SITE 1709 1709
 SQ SEQUENCE 2193 AA; 243209 MW; 04B3BC572A76E38 CRC64;
 Query Match
 Best Local Similarity 100.0%; Score 30; DB 1; Length 2193;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 EALFOG 6
 Db 1436 EALFOG 1441

RESULT 14
 POLG_EC09H STANDARD; PRT; 2193 AA.

AC O66849;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Genome polypeptide [Contains: Coat protein VP4 (P1A); Coat protein VP2 (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Picornain 2A (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein P3A; Genome-linked protein VP6 (P3B); Picornain 3C (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
 DE Echovirus 9 (strain H11).
 OS Enterovirus; ssRNA positive-strand viruses, no DNA stage; Picornaviridae; Enterovirus.

NCBI_TaxID=103915;
 RN (1)
 RP SEQUENCE FROM N.A.

RA MEDLINE=96435007; Pubmed=8837893;
 RA Zimmermann H., Eggers H.J., Krus W., Nielsen-Salz B.;

RT "Complete nucleotide sequence and biological properties of an infectious clone of prototype echovirus 9."

RL Virus Res. 39:311-319(1995).

-1- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE AT CERTAIN O/G SITES IN THE POLYPEPTIDE. THEY ARE CYSTEINE PROTEASES.

-1- CATALYTIC ACTIVITY: Selective cleavage of Gln-1-gly bond in the poliovirus polypeptide. In other picornavirus reactions Glu may be substituted for Gln, and Ser or Thr for Gly.

-1- CATALYTIC ACTIVITY: Selective cleavage of Tyr-1-gly bond in the poliovirus polypeptide. In other picornavirus reactions Glu may be substituted for Gln, and Ser or Thr for Gly.

-1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate + (RNA)(N).

-1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS, EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.

-1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

CC EMBL; U05876; AA50478.1; -
 DR HSSP; P03300; IPOV.
 DR MEROPS; C03.022; -
 DR MEROPS; C03.024; -
 DR InterPro; IPR000199; Cys-protease-3C.
 DR InterPro; IPR00138; Pico_P1A.
 DR InterPro; IPR000081; Pico_P2B.
 DR InterPro; IPR002527; Pico_P2B.
 DR InterPro; IPR000605; RNA_helicase.
 DR InterPro; IPR001205; RNA_pol_P3D.
 DR InterPro; IPR001676; RNV.
 DR Pfam; PF00073; Hvi_3.
 DR Pfam; PF00548; Cys-protease-3c; 1.

CC CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC; VP1/VP2 IS
 CC CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
 CC -1- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X84981; CAA59341.1; -
 DR HSP: G91734; 1EVI.
 DR InterPro: IPR004004; Calic1_pol_hel.
 DR InterPro: IPR000199; Cys-protease-3C.
 DR InterPro: IPR003138; Pico_P2A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR000605; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR InterPro: IPR001676; Rhv.
 DR Pfam: PF00073; rhv; 3.
 DR Pfam: PF00548; Cys-protease-3C; 1.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
 DR Pfam: PF00910; RNA_helicase; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR PRINTS: PR00918; CALICVIRUSNS.
 DR PRODOM: PD001125; Cys-protease-3C; 1.
 DR PRODOM: PD001274; Pico_P2B; 1.
 DR PRODOM: PD001306; Pico_P2A; 1.
 KM Polyprotein; Coat protein; Core protein; Transferrase; Myristate;
 KM RNA-directed RNA polymerase; Hydrolyase; Thiol protease.
 FT CHAIN 2 69
 FT CHAIN 70 330
 FT CHAIN 331 569
 FT CHAIN 570 859
 FT CHAIN 860 1009
 FT CHAIN 1010 1108
 FT CHAIN 1109 1437
 FT CHAIN 1438 1526
 FT CHAIN 1527 1548
 FT CHAIN 1549 1731
 FT CHAIN 1732 2193
 FT LIPID 2 2
 FT ACT_SITE 1695 1695
 FT ACT_SITE 1709 1709
 SQ SEQUENCE 2193 AA; 245056 MW; E6AD03091293C64A CXC64;
 Query Match 100.0%; Score 30; DB 1; Length 2193;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OS Echovirus 12 (strain Travis).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.
 OX NCBI_TaxId=103909;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Wild type;
 RX MEDLINE=95364006; PubMed=7637032;
 RA Kraus W., Zimmermann H., Zimmermann A., Eggers H.J., Nielsen-Salz B.;
 RT Infectious cDNA clones of echovirus 12 and a variant resistant
 RT against the uncoating inhibitor rhodamine differ in seven amino
 RT acids.
 RU J. Virol. 69:5853-5858 (1995).
 CC -1- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE
 CC AT CERTAIN O/G SITES IN THE POLYPROTEIN. THEY ARE CYSTEINE
 CC PROTEASES.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Gln-I-Gly bond in the
 CC poliovirus polyprotein. In other picornavirus reactions Gln may be
 CC substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of Tyr-I-Gly bond in the
 CC picornavirus polyprotein. In other picornavirus reactions Gln may
 CC be substituted for Gln, and Ser or Thr for Gly.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate +
 CC [RNA](N).
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC; VP1/VP2 IS
 CC CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
 CC -1- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X79047; CAA55650.1; -
 DR EMBL: X77708; CAA54785.1; -
 DR HSP: P21404; IDAM.
 DR MEROPS: C03.011; -
 DR MEROPS: C03.022; -
 DR InterPro: IPR004004; Calic1_pol_hel.
 DR InterPro: IPR000199; Cys-protease-3C.
 DR InterPro: IPR003138; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR000605; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR InterPro: IPR001676; Rhv.
 DR Pfam: PF00073; rhv; 3.
 DR Pfam: PF00548; Cys-protease-3C; 1.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
 DR Pfam: PF00910; RNA_helicase; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR PRINTS: PR00918; CALICVIRUSNS.
 DR PRODOM: PD001125; Cys-protease-3C; 1.
 DR PRODOM: PD001274; Pico_P2B; 1.
 DR PRODOM: PD001306; Pico_P2A; 1.
 KM Polyprotein; Coat protein; Core protein; Transferrase; Myristate;
 KM RNA-directed RNA polymerase; Hydrolyase; Thiol protease.
 FT CHAIN 2 69
 FT CHAIN 70 330
 FT CHAIN 331 568
 FT CHAIN 569 859
 FT CHAIN 860 1009
 FT CHAIN 1010 1108
 CORE PROTEIN P2B.

FT CHAIN 1109 1437 CORE PROTEIN P2C.
FT CHAIN 1438 1526 CORE PROTEIN P3A.
FT CHAIN 1527 1548 GENOME-LINKED PROTEIN VPg.
FT CHAIN 1549 1731 PICORNAIN 3C.
FT CHAIN 1732 2193 RNA-DIRECTED RNA POLYMERASE.
FT LIPID 2 MYRISTATE (BY SIMILARITY).
FT ACT_SITE 1695 1695 PROTEASE 3C (POTENTIAL).
FT ACT_SITE 1709 1709 PROTEASE 3C (POTENTIAL).
FT VARIANT 223 223 H -> Y (IN RHODAMINE-RESISTANT VARIANT).
FT VARIANT 228 228 G -> S (IN RHODAMINE-RESISTANT VARIANT).
FT VARIANT 376 376 I -> M (IN RHODAMINE-RESISTANT VARIANT).
FT VARIANT 643 643 Y -> C (IN RHODAMINE-RESISTANT VARIANT).
FT VARIANT 669 669 V -> A (IN RHODAMINE-RESISTANT VARIANT).
FT VARIANT 725 725 V -> A (IN RHODAMINE-RESISTANT VARIANT).
FT VARIANT 2094 2094 C -> R (IN RHODAMINE-RESISTANT VARIANT).
SQ SEQUENCE 2193 AA; 244803 MW; 1EA34E298F55130C CRC64;

Query Match 100.0%; Score 30; DB 1; Length 2193;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFQG 6
Db 1433 EALFQG 1438

Search completed: May 20, 2003, 11:35:01
Job time : 2.14163 secs


```

Q9RNM8
ID Q9RNM8 PRELIMINARY: PRT: 310 AA.
AC Q9RNM8:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE UDP-N-acetylglucosaminylglucosamine reductase.
GN MURB.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Zymomonas.
OX NCBI_TaxID=542;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ZM4;
RA Um H.W., Kang H.S.;
RT "Sequence analysis of 41E10 fosmid clone of Zymomonas mobilis.";
Submitted (AUG-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF119611; AAD53934.1; -.
DR InterPro: IPR001570; MurB.
DR InterPro: IPR001575; Oxid_FAD_bind.
DR Pfam: PF01565; FAD-binding_4; 1.
DR Pfam: PF02873; MurB_C; 1.
SQ SEQUENCE 310 AA; 33905 MW; 7DECA8C285C06B36 CRC64;

Query Match
Best Local Similarity 100.0%; Score 30; DB 2; Length 310;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
Db 190 EALFOG 195

RESULT 3
Q93VM4 PRELIMINARY: PRT: 522 AA.
AC Q93VM4:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Putative protease.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Pham P.K., Banh J., Banno F., Dale J.M.,
RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamuya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B.,
RA Lin J., Meyers W.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full length cDNA of gene T8H10.70/AT3957470 (GI:6706418).";
Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C., Toriumi M., Yu G.,
RA Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,
RA Jones T., Kamuya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E.,
RA Lam B., Lin J., Meyers W.C., Miranda M., Narusaka M., Nguyen M.,
RA Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full length cDNA of gene T8H10.70/AT3957470 (GI:6706418).";
Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AY035059; AAK59564.1; -.
DR InterPro: IPR001431; Peptidase_M16.

```

```

DR InterPro: IPR000169; SHprot_acsite.
DR Pfam: PF00675; Peptidase_M16; 1.
DR PROSITE: PS00143; INSULINASE; UNKNOWN.1.
DR PROSITE: PS00639; THIOI_PROTEASE_HIS; UNKNOWN.1.
SQ SEQUENCE 522 AA; 59556 MW; 81836BF3CE011A2 CRC64;

Query Match
Best Local Similarity 100.0%; Score 30; DB 10; Length 522;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
Db 235 EALFOG 240

RESULT 4
Q85082 PRELIMINARY: PRT: 757 AA.
AC Q85082:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Polypeptide precursor (Fragment).
OS Human poliovirus 1.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Enterovirus.
OX NCBI_TaxID=12080;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81273079; PubMed=6267593;
RA Semler B.L., Anderson C.W., Kitamura N., Rothberg P.G., Wisnart W.L.,
RA Wimmer E.;
RT "Poliovirus replication proteins: RNA sequence encoding P3-1b and the
RT sites of proteolytic processing."
Proc. Natl. Acad. Sci. U.S.A. 78:3464-3468(1981).
DR EMBL: M17494; AAA46911.1; -.
DR InterPro: IPR000199; Cys-protease-3C.
DR InterPro: IPR001205; RNA_pol_P3D.
DR Pfam: PF00548; Cys-protease-3C; 1.
DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR ProDom: PD001125; Cys-protease-3C; 1.
FT NON_TER 1 1
FT CHAIN 6 92 POTENTIAL.
FT CHAIN 93 114 POTENTIAL.
FT CHAIN 115 296 POTENTIAL.
FT CHAIN 297 757 POTENTIAL.
SQ SEQUENCE 757 AA; 84827 MW; C1B86D48A9D49060 CRC64;

Query Match
Best Local Similarity 100.0%; Score 30; DB 12; Length 757;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
Db 1 EALFOG 6

RESULT 5
Q9SCM5 PRELIMINARY: PRT: 989 AA.
AC Q9SCM5:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Protease-like protein.
GN T8H10.70.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]

```

RP SEQUENCE FROM N.A.
 RA Benes V., Rechmann S., Borkova D., Ansgorge W., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X., Quelet F., Salanoubat M.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL133248; CAB66104.1; -;
 DR InterPro: IPR001431; Peptidase_M16.
 DR InterPro: IPR000169; SHProl_acsite.
 DR Pfam: PF00675; Peptidase_M16; 1.
 DR PROSITE: PS00143; INSULINASE; 1.
 DR PROSITE: PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
 DR SEQUENCE 989 AA; 113070 MW; E12CF684FE14E96 CRC64;

Query Match 100.0%; Score 30; DB 10; Length 989;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 DB 235 EALFOG 240

RESULT 6
 O8TFN4 PRELIMINARY; PRT; 1866 AA.
 AC O8TFN4;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Chitin synthase C.
 GN CHSC.
 OS Colletotrichum graminicola (Anthracnose fungus) (Glomerella
 graminicola).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetes Incertae sedis; Phyllachorales; Phyllachoraceae;
 OC Glomerella.
 OX NCBI_Taxid=31870;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Werner S., Deising H.B.;
 RT "A csm-like chitin synthase of the maize pathogen Colletotrichum
 graminicola."
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY053447; AAL23719.1; -;
 DR SEQUENCE 1866 AA; 207182 MW; 9126E926330DE1BF CRC64;

Query Match 100.0%; Score 30; DB 3; Length 1866;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 DB 1145 EALFOG 1150

RESULT 7
 O91QW1 PRELIMINARY; PRT; 1930 AA.
 AC O91QW1;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Polyprotein (Fragment).
 OS Human echovirus 30.
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.
 OX NCBI_Taxid=41846;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DR/ROMA97;

RA Muscillo M., La Rosa G., Marianeili C., Capobianchi M.R.,
 RA Zaniratti S., Cattani P., Manzara S., Fadda G., Comparcola D.,
 RA Tacca F.;
 RT "A phylogenetic analysis of the two echovirus 30 isolated in Rome
 (Italy) in 1997 from an outbreak of aseptic meningitis associated with
 swimming pools."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ295172; CAC39011.1; -;
 DR MEROPS: C03.011; -;
 DR InterPro: IPR000199; Cys-protease-3C.
 DR InterPro: IPR003138; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR001676; Rhv.
 DR InterPro: IPR001205; RNA_helicase.
 DR InterPro: IPR000605; RNA_helicase.
 DR Pfam: PF00548; Cys-protease-3C; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF00073; rhv; 3.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
 DR Pfam: PF00910; RNA_helicase; 1.
 DR ProDom: PD001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 FT CHAIN 1
 FT CHAIN 70
 FT CHAIN 331
 FT CHAIN 369
 FT CHAIN 864
 FT CHAIN 1000
 FT CHAIN 1099
 FT CHAIN 1439
 FT CHAIN 1528
 FT CHAIN 1550
 FT CHAIN 1733
 FT CHAIN 1930
 FT NON_TER 1930
 SQ SEQUENCE 1930 AA; 214968 MW; 559CFD7CC4F52FB2 CRC64;

Query Match 100.0%; Score 30; DB 12; Length 1930;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 DB 1434 EALFOG 1439

RESULT 8
 O8VOC4 PRELIMINARY; PRT; 2063 AA.
 AC O8VOC4;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Polyprotein (Fragment).
 OS Human poliovirus 1.
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.
 OX NCBI_Taxid=12080;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3788ALB96;
 RX MEDLINE=21635047; PubMed=11773144;
 RA Marturano J., Fiore L.,
 RT "Investigation of the Presence of Recombinant Polioviruses in the Hit
 Population in Albania during the 1996 Outbreak."
 RL J. Clin. Microbiol. 40:316-317(2002).
 DR EMBL: AY056701; AAL33795.1; -;
 DR InterPro: IPR000199; Cys-protease-3C.
 DR InterPro: IPR000345; CysC_heme_bind.

FT	CHAIN	340	>825	VP3.
----	-------	-----	------	------

```

FT CHAIN 826 >1137 VP1.
FT CHAIN 1128 >1276 PROTEINASE P2-A.
FT CHAIN 1277 >1403 P2-B.
FT CHAIN 1404 >1469 P2-C.
FT CHAIN 1670 >1789 P3-A.
FT CHAIN 1790 >1811 VP.
FT CHAIN 1812 >2001 PROTEINASE P3-C.
FT CHAIN 2002 >2063 POLYMERASE P3-D.
FT NON_TER 2063
SQ SEQUENCE 2063 AA: 229820 MW: 958FE9141805F2A1 CRC64;

Query Match
Best Local Similarity 100.0%; Score 30; DB 12; Length 2063;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
DB 1447 EALFOG 1452

RESULT 11
Q8QU26 PRELIMINARY; PRT; 2171 AA.
ID Q8QU26;
AC Q8QU26;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Polypeptide.
OS Porcine enterovirus 10.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Enterovirus.
OX NCBI_TaxID:106967;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-IP 54;
RA Krumbholz A., Dauber M., Henke A., Birch-Hirschfeld E., Knowles N.J.,
RA Steizner A., Zell R.;
RT "Sequencing of porcine enterovirus (PEV) groups II and III reveals
RT unique features of both virus groups."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF363455; AAM00276.1; -
SQ SEQUENCE 2171 AA: 240751 MW: 789FE0C3A7FD378 CRC64;

Query Match
Best Local Similarity 100.0%; Score 30; DB 12; Length 2171;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
DB 1412 EALFOG 1417

RESULT 12
Q9PYB7 PRELIMINARY; PRT; 2174 AA.
ID Q9PYB7;
AC Q9PYB7;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Polypeptide.
OS Bovine enterovirus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Enterovirus.
OX NCBI_TaxID:12064;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K2577;
RA MEDLINE-99429271; PubMed-10501163;
RA McCarthy F.M., Smith G.A., Mattick J.S.;
RT "Molecular characterisation of Australian bovine enteroviruses."
RL Vet. Microbiol. 68:71-81(1999).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-K2577;
RA McCarthy F.M., Smith G.A., Mattick J.S.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF123432; AAF05693.1; -
DR HSSP; P12915; 1BEV.
DR InterPro; IPR004004; Calic_pol_hel.
DR InterPro; IPR000199; Cys-protease-3C.
DR InterPro; IPR003138; Pico_P1A.
DR InterPro; IPR000081; Pico_P2A.
DR InterPro; IPR002527; Pico_P2B.
DR InterPro; IPR001676; Rhv.
DR InterPro; IPR000605; RNA_helicase.
DR InterPro; IPR001205; RNA_pol_P3D.
DR Pfam; PF00548; Cys-protease-3C; 1.
DR Pfam; PF02226; Pico_P1A; 1.
DR Pfam; PF00947; Pico_P2A; 1.
DR Pfam; PF01552; Pico_P2B; 1.
DR Pfam; PF0073; Rhv; 3.
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR Pfam; PF00910; RNA_helicase; 1.
DR PRINTS; PR00918; CALICVIRUSNS.
DR ProDom; PD001125; Cys-protease-3C; 1.
DR ProDom; PD001274; Pico_P2B; 1.
DR ProDom; PD001306; Pico_P2A; 1.
FT CHAIN 1
FT CHAIN 70 71 VP4.
FT CHAIN 318 559 VP2.
FT CHAIN 560 839 VP3.
FT CHAIN 840 989 VP1.
FT CHAIN 990 1088 PROTEINASE 2A.
FT CHAIN 1089 1418 2B.
FT CHAIN 1419 1507 2C.
FT CHAIN 1508 1530 3A.
FT CHAIN 1531 1713 VP.
FT CHAIN 1714 2174 PROTEINASE 3C.
SQ SEQUENCE 2174 AA: 242178 MW: 3D4787508B05288 CRC64;

Query Match
Best Local Similarity 100.0%; Score 30; DB 12; Length 2174;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EALFOG 6
DB 1414 EALFOG 1419

RESULT 13
Q9PYB6 PRELIMINARY; PRT; 2174 AA.
ID Q9PYB6;
AC Q9PYB6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Polypeptide.
OS Bovine enterovirus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Enterovirus.
OX NCBI_TaxID:12064;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SL305;
RA MEDLINE-99429271; PubMed-10501163;
RA McCarthy F.M., Smith G.A., Mattick J.S.;
RT "Molecular characterisation of Australian bovine enteroviruses."
RL Vet. Microbiol. 68:71-81(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SL305;
RA McCarthy F.M., Smith G.A., Mattick J.S.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF123433; AAF05694.1; -
DR HSSP; P12915; 1BEV.
DR InterPro; IPR004004; Calic_pol_hel.

```

DR InterPro: IPR000199; Cys-protease-3C.
 DR InterPro: IPR003138; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR001676; Rhv.
 DR InterPro: IPR000605; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR Pfam: PF00548; Cys-protease-3C; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF00073; Rhv; 3.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
 DR Pfam: PF00910; RNA_helicase; 1.
 DR PRINTS: PR00918; CALICIVIRUS.
 DR ProDom: PD001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 DR CHAIN 1 69 VP4.
 FT CHAIN 1 69 VP4.
 FT CHAIN 317 VP2.
 FT CHAIN 318 559 VP3.
 FT CHAIN 560 839 VP1.
 FT CHAIN 840 989 PROTEASE 2A.
 FT CHAIN 990 1088 2B.
 FT CHAIN 1089 1418 2C.
 FT CHAIN 1419 1507 3A.
 FT CHAIN 1508 1530 VPG 3B.
 FT CHAIN 1531 1713 PROTEASE 3C.
 FT CHAIN 1714 2174 POLYMERASE 3D.
 SO SEQUENCE 2174 AA; 242286 MW; 9AADC6487649A2BC CRC64;

Query Match Best Local Similarity 100.0%; Score 30; DB 12; Length 2174;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 1414 EALFOG 1419

RESULT 14
 O8V639 PRELIMINARY; PRT; 2183 AA.
 AC O8V639;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Polypeptide.
 OS Human coxsackievirus B4.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.
 NC NCB1_TaxID=12073;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=E2 VARIANT;
 RA Lindberg A.M., Andersson P.;
 RT "Molecular analysis of a coxsackievirus B4 strain E2 variant."
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF131939; AAL37156.1; -
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR000199; Cys-protease-3C.
 DR InterPro: IPR003138; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR001676; Rhv.
 DR InterPro: IPR000605; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR Pfam: PF00548; Cys-protease-3C; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF00073; Rhv; 3.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.

DR Pfam: PF00910; RNA_helicase; 1.
 DR ProDom: PD001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 DR SMART: SM00382; AAA; 1.
 SO SEQUENCE 2183 AA; 244045 MW; BC34D472E37953D CRC64;

Query Match Best Local Similarity 100.0%; Score 30; DB 12; Length 2183;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 Db 1423 EALFOG 1428

RESULT 15
 O9OL88 PRELIMINARY; PRT; 2184 AA.
 AC O9OL88;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Polypeptide.
 OS Human coxsackievirus B6.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Enterovirus.
 NC NCB1_TaxID=74561;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SCHMITT;
 RX MEDLINE-99432326; PubMed-10500285;
 RA Martino T.A., Tellier R., Petric M., Irvin D.M., Afshar A., Liu P.P.;
 RT "The complete consensus sequence of coxsackievirus B6 and generation
 of infectious clones by long RT-PCR."
 RL Virus Res. 64:77-86(1999).
 DR [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SCHMITT;
 RA Martino T.A., Tellier R., Petric M., Irvin D.M., Afshar A., Liu P.P.;
 RL Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF105342; AAF12719.1; -
 DR HSP: P21404; 1DM.
 DR MEROPS: C03.022; -
 DR InterPro: IPR004004; Calic1_pol_hel.
 DR InterPro: IPR000199; Cys-protease-3C.
 DR InterPro: IPR003138; Pico_P1A.
 DR InterPro: IPR000081; Pico_P2A.
 DR InterPro: IPR002527; Pico_P2B.
 DR InterPro: IPR001676; Rhv.
 DR InterPro: IPR000605; RNA_helicase.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR Pfam: PF00548; Cys-protease-3C; 1.
 DR Pfam: PF02226; Pico_P1A; 1.
 DR Pfam: PF00947; Pico_P2A; 1.
 DR Pfam: PF01552; Pico_P2B; 1.
 DR Pfam: PF00073; Rhv; 3.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
 DR Pfam: PF00910; RNA_helicase; 1.
 DR PRINTS: PR00918; CALICIVIRUS.
 DR ProDom: PD001125; Cys-protease-3C; 1.
 DR ProDom: PD001274; Pico_P2B; 1.
 DR ProDom: PD001306; Pico_P2A; 1.
 DR CHAIN 1 69 VP4.
 FT CHAIN 1 69 VP4.
 FT CHAIN 70 330 VP2.
 FT CHAIN 331 568 VP3.
 FT CHAIN 569 850 VP1.
 FT CHAIN 851 1000 P2-A.
 FT CHAIN 1001 1099 P2-B.
 FT CHAIN 1100 1428 P2-C.
 FT CHAIN 1429 1517 P3-A.
 FT CHAIN 1518 1539 P3-B.
 FT CHAIN 1540 1722 P3-C.

FT CHAIN 1723 2184 P3-D, 48B8268B1C523AD8 CRC64;
 SO SEQUENCE 2184 AA; 243253 MM; 48B8268B1C523AD8 CRC64;

Query Match 100.0%; Score 30; DB 12; Length 2184;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EALFOG 6
 11111
 Db 1424 EALFOG 1429

Search completed: May 20, 2003, 11:37:49
 Job time : 2.41398 secs